PLayful Environment for Inclusive leArning Design in Europe

Intellectual Output 4: Evaluation kit for inclusion-oriented collaborative learning activities

“PLayful Environment for Inclusive leArning Design in Europe”

“PLEIADE”

Project No. 2020-1-IT02-KA201-080089

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

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Project ref. number	2020-1-IT02-KA201-080089
Project title	**PLEIADE - PLayful Environment for Inclusive leArning Design in Europe**
Document title	Evaluation kit for inclusion-oriented collaborative learning activities (version 2.1)
Document Type	Intellectual Output [accompanying document]
Document version	2.1
Previous version(s)	1.0 (March 11, 2022) 1.1 (June 8, 2023) 2.0 (July 7, 2023)
Planned date of delivery	June 31 2023
Language	English
Dissemination level	Public
Number of pages	117
Partner responsible	Sofia University
Author(s)	Nikolina Nikolova (Sofia University), Pencho Mihnev (Sofia University), Temenuzhka Zafirova-Malcheva (Sofia University), Ralitza Stamenkova (Sofia University), Stanislav Ivanov (Sofia University), Donatella Persico (CNR-ITD), Marcello Passarelli (CNR-ITD), Francesca Pozzi (CNR-ITD), Erica Volta (CNR-ITD)
With contributions by:	Gergova, A.; Lonigro, M.; Tsairi, K.; Kyriakides, K.
Revised by:	Donatella Persico (CNR-ITD), Tsairi Kalliopi (Avgoulea–Linardatou School)
Abstract	This document is intended as a text file annexed to IO4 and is distributed as an accompanying document to the PEIADE Evaluation kit, which is available through: https://pleiade-project.eu/outputs . The main purpose of this text file is to document in detail the process of the Evaluation kit development, the intermediate and supporting products, and to provide scientific reasoning of its validity and reliability. The document describes the main activities carried out during the Evaluation kit development

	and the responsibilities taken by the partners. It provides a description of the developed Evaluation kit and use-cases, supporting its usage by external users in their attempts to develop and enact inclusive collaborative learning designs.
Keywords	Evaluation kit, criteria for inclusiveness, inclusive potential, inclusive power
DOI	https://doi.org/10.60063/NN.2023.0089.95
How to cite[1]	Nikolova, N., Mihnev, P., Zafirova-Malcheva, T., Stamenkova, R., Ivanov, S., Persico, D., Passarelli, M., Pozzi, F., Volta, E. (2023). *Evaluation kit for inclusion-oriented collaborative learning activities* – PLEIADE Intellectual Output 4. https://doi.org/10.60063/NN.2023.0089.95

[1] Use APA style

Internal project peer review process

Reviewed by	**Role, Organisation**	**Date**
Reviewer 1	Donatella Persico	18/07/23
Reviewer 2	Tsairi Kalliopi	15/07/23
Reviewer 3[2]	–	–
Approved by	Francesca Pozzi	<24/07/2023>
Francesca Pozzi	Project Manager	<24/07/2023>
Marcello Passarelli	Quality Manager	<24/07/2023>

[2] Optional

Table of Content

List of Figures

List of Tables

1 Executive summary

This document is intended as a text file accompanying the Intellectual Output 4 (IO4) of Erasmus+ project “PLayful Environment for Inclusive leArning Design in Europe” (PLEIADE), the Evaluation kit for inclusion-oriented collaborative learning activities. According to the project proposal, the Evaluation kit purpose is to assess the designs produced by the teachers (by means of their **inclusive potential**) and select good practices examples (by means of their **inclusive power**), as well as to evaluate the project results. Additionally, it is intended to support schools, teachers and school leaders beyond project activities in self-assessing the inclusive affordances of their activities.

The main purpose of this text file is to document in detail the process of the Evaluation kit development, the intermediate and supporting products, and to provide scientific reasoning of its validity and reliability. The document describes the main activities carried out during the Evaluation kit development and the responsibilities taken by the partners involved. It provides a description of developed kit and use-cases, supporting its usage by external parties in their attempts to develop and enact inclusive collaborative learning designs.

The first kit prototype was delivered in February 2023 (M30), and the revised and extended version was available in May 2023 (M33). The PLEIADE teachers tested the current version for validity (usability, clear meaning, ease of use). Its reliability is measured by the inter-raters’ agreement and by Cronbach Alpha coefficient.

The current version of the Evaluation kit is accessible through the PLEIADE website: https://pleiade-project.eu/outputs. It is primarily in English, but also localized in Italian, Bulgarian and Greek language. Accompanying documents – User Guide, validation tools, are also accessible through the website.

2 Introduction

The dynamic political and economic world-wide context leads to quick changes of the local educational environment in each country. Migration of people and pandemic situation poses challenges, accompanied by the threat of excluding different groups of pupils from effective educational process. Students, presenting refugee’s and minority families often have problems with the native language of the host country (Morrice, 2019). Students with low social status or low level of digital skills could not participate at an equal level in the online learning process (Davidson & Carr, 2010). Students with learning and / or physical disabilities, and other special educational needs are disconnected from their teachers and school facilitators (De Leeuw, De Boer, & Minnaert, 2019).

The presented problems are addressed by the PLEIADE Erasmus Plus K201 project (https://pleiade-project.eu). It supports the efforts of the teachers to integrate systematically inclusive practices by focusing on their Learning Design competences (Pleiade project, 2021). Teachers from Bulgaria, Italy, Greece and Cyprus accepted the challenge to pass through Blended Learning Activities (BTA) (Passarelli, Dagnino, Persico, Pozzi, & Nikolova, 2021) lasting 13 months and, during and after the training, to develop and implement high quality inclusive learning designs, based on collaborative learning approaches, to be shared as open educational resources (OER).

The process of development and selection of the most effective designs requires development of a criteria matrix (or rubric) for the evaluation of the designs's **inclusive potential**. It should contain the main criteria and relative indicators ensuring the **potential** (at the design level) and the **power** (at the enactment level) of inclusive teaching and learning, taking into account the various factors influencing the process of exclusion from the learning process. It should support pointing out weaknesses and possible improvements of the learning designs and their enactments.

This structured rubric has been developed by the PLEIADE project partners and incorporated into an evaluation kit, reusable by stakeholders outside the project to further serve a larger target of European teachers and schools, looking for social inclusion of their students through collaborative activities, integrated into the learning designs and managed by the teachers during enactments.

2.1 Aim of this document

This document is intended as part of IO4 and is distributed as release notes to the PLEIADE Evaluation kit. Particularly, two are the expected results from IO4:

- An interactive resource – the Evaluation kit
- An accompanying text file – which is this document.

The main aim of the Evaluation kit for inclusion-oriented collaborative learning activities is to assess:

- the inclusive potential of a design, by checking its features against a number of indicators for inclusive design (before the intervention):
- the actual inclusive power of the teaching intervention, by collecting data about the students’ experience, during the intervention and following it.

The main scope of this text file is to document in detail the design and process of the development of evidence-based evaluation kit that provide a scientifically grounded set of indicators, intended to operationalize the assessment of a design (in terms of inclusive potential) and of a teaching intervention (in terms of its inclusive power).

According to the project proposal, the evaluation kit should be represented as an interactive resource, easily understood and manageable, clear in meaning and widely shareable in the community of student teachers.

The kit is developed in English and localized in Italian, Bulgarian and Greek language.

In line with the revised Gantt chart, IO4 started at M8, in parallel with the BTA during the Training phase. The first part of the document was written at M20, providing a base for the First draft of the Delphi study and for supporting development of the high level of inclusive potential designs by the PLEIADE teachers at the end of the training phase. The second part – description of the Delphi study process and intermediate result was written at M31, and completed and revised at M32 (June 2023). The complete second version of the document was released at the beginning of M33 (July 03, 2023) and internally revised by the end of July 2013.

The document is accompanied by appendices, containing draft lists of criteria, data, validation tools, QR codes for reusability of the kit, and User guide.

2.2 Structure of this document

The document is structured as follow:

- Section 3. provides a general introduction to IO4.
- Section 4. describes methodology of the development of the Evaluation kit. The aims and the main stages of the process are presented as well as the interaction between the Delphi study intermediate results and the main

contributors – the experts from CNR-ITD and UniSofia, external experts’ focus group, and PLEIADE teachers.

- Section **5.** describes the process of extracting the main inclusive criteria through literature review and through participatory practices – peer review of the teachers’ first collaborative designs and teachers’ contribution through Nominal Group Technique. The section provides a scientifically reasoned ground for further development and refinement of the list of criteria and sub-criteria trough Delphi study.
- Section 6. reveals details about the Delphi study procedure. It publishes the intermediate list produced, which is a base for the development of the prototype of the kit.
- Section 7. describes the Evaluation kit in two versions: the Evaluation kit for individual user and Evaluation kit for multiple users are described in details. Use-cases for both of them are provided.
- Section 8. presents the validation methodology, providing reasoning of methods used. It describes the validation process in its complexity – interweaving and interacting with Delphi study and the Evaluation kit development. The section provides evidence for usefulness and reliability of the Evaluation kit.
- The conclusions and future prospects are presented in Section 9.

3 Description of the IO4

The Intellectual Output 4 consists of an interactive Evaluation kit for measuring the inclusive potential of a collaborative learning design and the inclusive power of its enactment. It is developed through literature review and Delphi study participatory approach. The interactivity is provided by Google sheets and Google forms. The Evaluation kit consists of two tools – for individual user and for multiple users. The kit usefulness is validated in three stages during its development. The inter-raters reliability and the list of subcriteria reliability are calculated. The coefficients ensure the reliability of the whole kit.

In the context of the PLEIADE project, the Evaluation kit ensured development of collaborative learning designs with higher level of inclusiveness and good enactment practices.

For external users the kit could serve many cases of in-service and pre-service teachers training – self-assessment of the inclusiveness of a design or practice, peer assessment during teachers’ training, monitoring process by external authorities and others.

In terms of sustainability and transferability, the Evaluation kit and accompanying documents – User Guide, list of criteria and sub-criteria, validation tools and method will remain available on the PLEADE website (https://pleiade-project.eu/) for (at least) 24 months after the project ends.

The production of IO4 intertwined with IO5. The following specific tasks were accomplished during the project lifespan:

1. Literature review of state of the art research on recommendations for the design and enactment of inclusive collaborative learning and assessment.
2. Production of a draft set of indicators and strategies to assess learning designs (indicators for inclusive design practice), based on the above literature review.
3. Production of a draft set of indicators and strategies to assess learning interventions (indicators for inclusive teaching practice) based on the above literature review.
4. Revision of the above indicators and strategies based on a participatory approach (focus groups, Delphi Study, Nominal Group technique) involving PLEIADE teachers and other stakeholders.
5. Development of the prototype kit based on revised indicators.
6. Quali-quantitative testing of the prototype kit on the designs, produced by the PLEIADE teachers: the testing is aimed to assess the prototype perceived ease

of use and usability of the kit, and to ensure good raters’ agreement for untrained users of the kit. This task was carried out in synergy with the “assessment of the PLEIADE teaching interventions” task involved in [O5] development.

7. Prototype revision based on test outcomes and development of the final kit, consisting of a set of interactive and printed checklists guiding the assessment process.

3.1 Timeline

Originally, the IO4 should start at M18 (February 2022). By a bilateral agreement between CNR-ITD and UniSofia, the output start was shifted to the M8 (April 2021) in order to deliver the literature review in advance and to provide a base for the Delphi study and for on-going assessment of the PLEIADE collaborative designs. The intellectual output development lasted until M34 (June 23).

The main tasks were distributed between UniSofia and CNR-ITD.

CNR-ITD was responsible for the development of the list of criteria and sub-criteria managing the Delphi study, while UniSofia developed the technical presentation of the interactive tool and managed the procedure of validation of the tool (Figure 1).

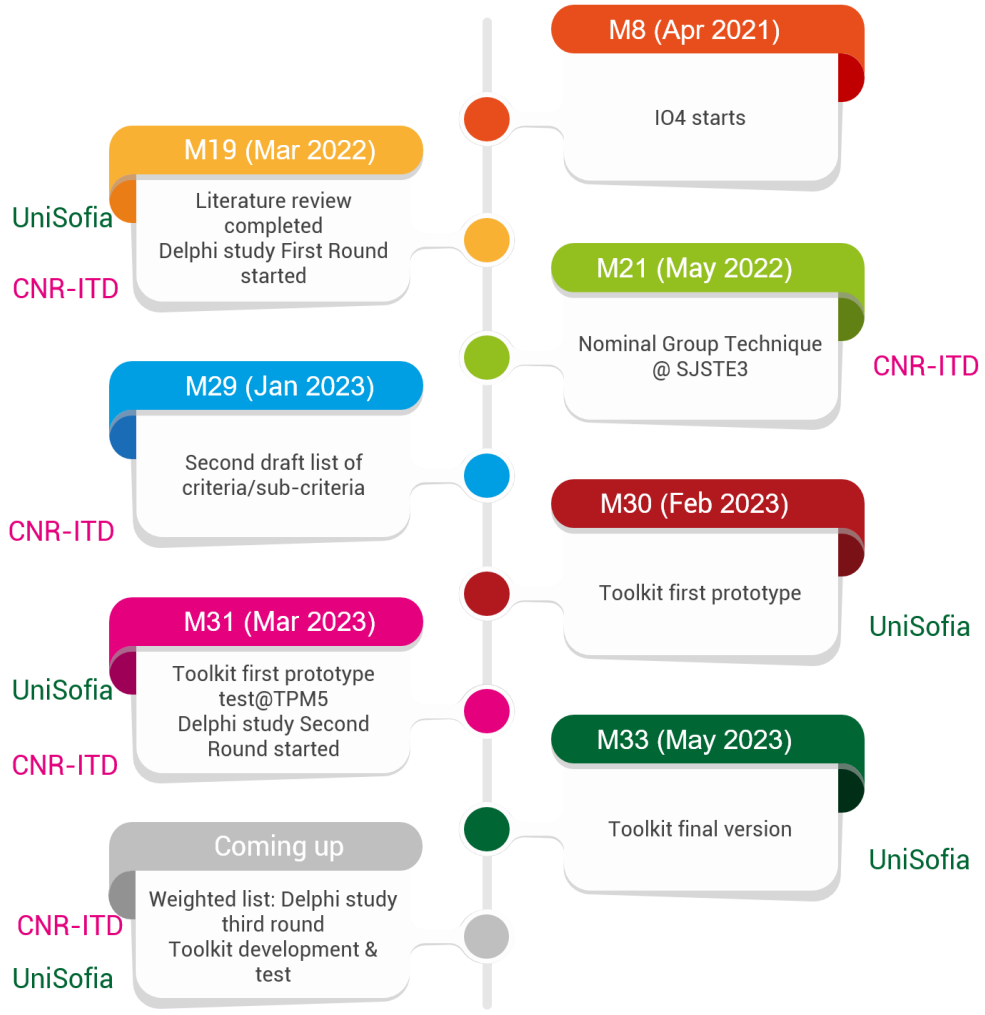


Figure 1. Timeline of the IO4 development

3.2 Partners' contribution

The partners involved with major effort (in terms of working days) in this output were CNR-ITD and University of Sofia, as follows:

- CNR-ITD led the Delphi study and development of the clear and concise list of operationalized criteria and subcriteria, providing meaningful guidance for development on collaborative learning design with high level of inclusiveness and for its enactment in terms of ensuring the inclusiveness throughout the teacher's intervention
- CNR-ITD designed and implemented the Nominal Group Technique as a participatory approach for refining the initial criteria list.
- UniSofia performed initial literature review and analysis of the experts' and peers' reviews of the initial PLEIADE designs, both providing an evidence-based ground for the criteria list development.
- UniSofia is responsible also for technical development of the kit and testing its ease of use, usefulness and good raters' agreement.

The partnering schools were involved with less effort (in terms of working days), contributing to the following tasks:

- Providing data for the criteria list through the peer review of the initial designs.
- Providing data for the criteria list through the Nominal Group Technique.
- Participation in all stages of testing the Evaluation kit
- Translation of the kit into Italian, Bulgarian and Greek language.

4 Methodology for development a kit for the evaluation of the inclusive potential and inclusive power of a learning design / enactment

4.1 Aims

The Evaluation kit for inclusion-oriented collaborative learning activities is used to assess the designs produced by the teachers, and select good practice examples, as well as to evaluate the project results. Additionally, it is intended to support schools, teachers and school leaders beyond project activities in self-assessing the inclusive affordances of their activities.

This intellectual output takes on board state of the art research results concerning inclusion and the experience gained in the project in order to develop a kit for assessing:

- the **inclusive potential** of a design, by checking its features against a number of indicators for inclusive design (before the intervention):
- the actual **inclusive power** of the teaching intervention, by collecting data about the students’ experience, during the intervention and following it.

Research evidence and guidelines for practice concerning effective inclusive teaching is distilled to produce a **clear, concise, and grounded set of draft indicators**, intended to operationalise the assessment of a design (in terms of **inclusive potential**) and a teaching intervention (in terms of its **inclusive power**).

4.2 Main stages

The Evaluation kit development includes the following main stages:

- Literature review: extracting main criteria.
- Empirical research on PLEIADE designs development: mapping the peer review to the literature review, leading to refined groups of criteria (main criteria and sub-criteria).
- Impact on IO1 & IO5 providing directions for following the criteria for development and implementation of inclusive learning designs.
- Delphi study for development of a structured list of criteria and two operationalized lists of sub-criteria (for measuring inclusive potential and inclusive power), which to be incorporated in IO4 Evaluation kit.
- Development of the prototype kit based on revised indicators.
- Validation of the prototype of the evaluation kit by quali-quantitative testing on the designs, produced by the PLEIADE teachers.

- Prototype revision, based on test outcomes.
- Final kit development.

Figure 2 illustrates the whole process of IO4 development.

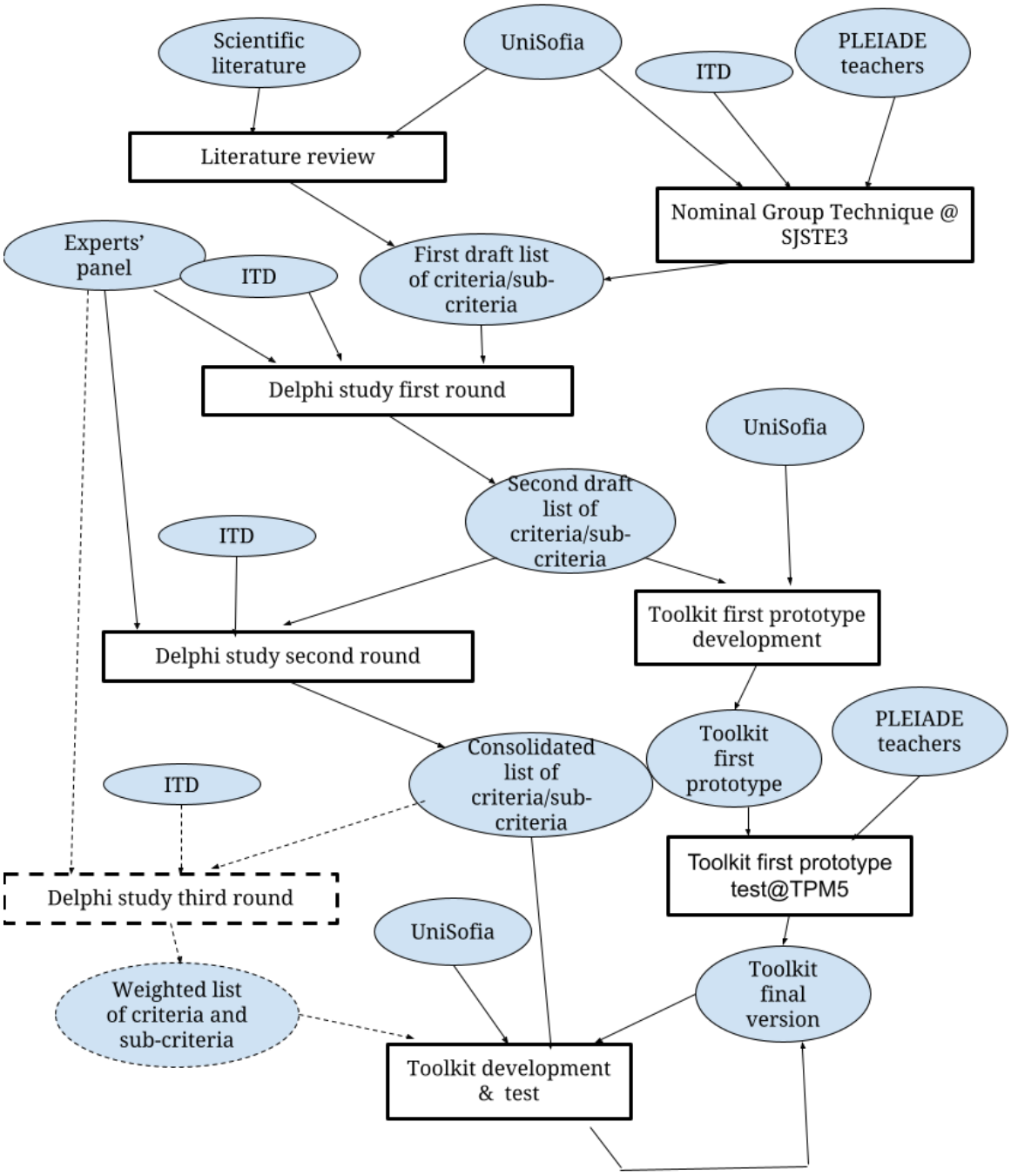


Figure 2. The Evaluation kit development methodology[3] (dotted lines represent activities that will be carried out after the end of the project)

[3] In this picture, Petri Nets (a formalism used in computer science to represent parallel processes/activities and their concurrent use of input and output resources) are used with some adaptations, to represent the workflow in IO4 development. In this formalism, rectangular shapes represent activities while ovals represent input/output resources, including actors involved (Olimpo, 2011).

A literature review on inclusive learning is performed to extract the first set of important criteria. Analysing the scientific research on the topic, the criteria are grouped according to the main characteristics of a learning design – learning goals and objectives, learning resources, classroom management, learning techniques and activities, etc. Literature review on relevant indicators is also performed.

After forming a main set of criteria and indicators from theoretical aspects, the empirical study on the process of development of first inclusive designs by the PLEIADE teachers is taken place. The focus was on the peer-review of the designs as well as on the feedback provided by CNR-ITD and UniSofia experts. The qualitative feedback and reflection provide information on what the reviewers are looking for and what indicators they provide as a reasoning, even on an intuitive practical level. The approach allows to map the scientific theoretical results to the practice and to formulate relevant, understandable for the practitioner criteria and indicators for inclusiveness of the learning on the both levels – design and enactment. As a result, the first basic list of inclusive criteria is developed. It served both aims – 1) producing first draft list of criteria and operationalized sub-criteria for Delphi study, aiming to operationalize the evaluation in such a way that the list to provide also directions and feedback for achieving inclusive potential and inclusive power, and 2) to guide the PLEIADE designs development in order to guarantee their inclusive potential at the very beginning and to improve it during the designs development process.

Further, the Delphi study has run, and, in parallel, the intermediate results interacted with the last design development and enactments (IO4). The validation process has accompanied both, providing in-time information about the direction, need of corrections and improvements, etc.

5 Extracting main criteria groups for evaluation of inclusive potential and inclusive power

Having so different aspects of exclusion reasons, the project scientific team, represented by Institute for Educational Technology (CNR-ITD), Italy, and Sofia University, Bulgaria, tries to identify common criteria for evaluation of the inclusive potential of a design for collaborative learning (Zafirova-Malcheva, Antonova, Stamenkova, Nikolova, & Mihnev, 2022). The role of the criteria is to guide teachers in different contexts to design inclusive teaching processes, providing a base for equal access to a high level of education.

The process of extracting criteria was going on in parallel with blended teaching activities (BTAs) for teachers, provided in an environment of metaphor of space voyage journey (Passarelli, Dagnino, Persico, Pozzi, & Nikolova, 2021). During the BTAs, the teachers were invited to develop first drafts of their designs (two scenarios by school), focused on the inclusive potential of the designs. After that they participated in a peer review process, reflecting on how, where and why they see the inclusive potential of the scenarios, developed by their colleague. The objective was to comment and rate these scenarios based on the effect on children with cultural, linguistic and socio-economic difficulties.

The methodology of extracting criteria is organised in two stages. During the first stage, groups of criteria are extracted, following a thorough desktop research. During the second stage, the schools, supported by the CNR-ITD team, developed the first working set of learning scenarios. The data sources, provided by them, containing results of the peer-review, discussions in peers, and feedback, provided by experts from CNR-ITD, are analysed and the chosen factors are adjusted. Finally, the analysis of the discussions, comments and revisions are mapped to the literature research results. The last part summarises the criteria for evaluation of the inclusive potential of a learning scenario before its implementation in a real learning situation.

5.1 Literature Review for extracting criteria and indicators for evaluating inclusive potential of a learning design and inclusive power of an enactment

During the first stage of extracting criteria a literature review was made (Zafirova-Malcheva, Antonova, Stamenkova, Nikolova, & Mihnev, 2022). It identified several main sources, based on which the main criteria and indicators for inclusive potential of learning scenarios were derived.

Jordan and McGhie-Richmond, in their study "Identifying Effective Teaching Practices in Inclusive Classrooms" (Jordan & McGhie-Richmond, 2014), present an observation scale used to *rate effective inclusive instructional practices*, applicable for both as a self-rating and as a third-party measurement scale of effective teaching practices. Classroom observation system for inclusive teaching practices described by the authors contains six groups of inclusive indicators related to classroom and time management, lesson presentation, large group and whole class instruction scaffolding, small group and individual instruction, teaching style and classroom tone. Based on this scale, several important inclusion criteria were identified: in terms of content and activities, is important that they are designed so that “activate prior experiences and knowledge relevant to the topics, strategies or skills to be learned”, “forecast upcoming learning content”, “contains pre-set clear instructions, expected results and execution time”, “provides error correction procedures”, it must require explicitly the participation of all students; for assessment is important to “summarize accomplishments of individuals and group”; to assure accessibility the “rules must be applied equally for all”; the encouragement of students collaboration also is very important criteria; in terms of social and personal relationships two very important criteria have been identified – “focusing on student relationships” and “including rules that require respect and support for other students” (Jordan & McGhie-Richmond, 2014).

Loreman, Forlin and Sharma make a systematic literature review of measuring indicators of inclusive education (Loreman, Forlin, & Sharma, 2014). In this review, they outline 13 themes in the international literature that should be considered in the development of a set of indicators for measuring inclusive education. They also identify two frameworks for examining inclusive education, on the one hand, they use three levels – micro, meso and macro, to examine indicators for inclusive education, and on the other they apply the inputs-processes-outcomes model. Combining these two frameworks, they identify the themes and describe the criteria. The focuses are on conditions / climate and participation, which determines extremely important criteria that inclusive education must meet such as “collaboration and shared responsibility”, “remove barriers and inequalities”, “valued and respected all learners”, and their “active involvement”, build “friendships/relationships”, stimulating “contacts/interactions” and “membership of clubs, teams”, “peer support” and “acceptance by peers”, “parent involvement” and “meet individual needs” (Loreman, Forlin, & Sharma, 2014).

In its article “Criteria for Designing Blended Learning Materials for Inclusive Education: Perspectives of Teachers and Producers", Bosse (Bosse, 2015) is focused on the *design of learning materials for inclusive education*. The author considers the

problem from the perspectives of experienced teachers and from the perspectives of the producers. The research is divided into two modules. In the first one, based on teachers’ interviews, the author describes recommendations for the design of blended learning materials for inclusive education. Based on these recommendations in the second module the producers develop criteria for the design of materials, modules, and activities for inclusive education (Bosse, 2015). Although the criteria presented in the article are not directly aimed at students, they outline extremely important requirements for the learning content, its structure and accessibility, which is a critical element for certain learners’ groups. As basic principles for blended learning materials are indicated “individualisation and personalisation”, “cooperative/collaborative learning”, “activity-oriented instruction”, “web accessibility and universal design”. In the produced checklist are included important criteria such as “clear explanations”, “reference to the real world”, “content-related assistance”, “simple, intuitive operability (no lengthy instructions necessary)” (Bosse, 2015).

The “Index for inclusion: developing learning and participation in schools” (Booth & Ainscow, 2011) was developed for English schools by the Centre for Studies on Inclusive Education (CSIE) with the help of a team of teachers, parents, governors and researchers from the United Kingdom. Although the “Index” was produced for English schools, it has been adapted for use in many other countries and translated into thirty seven languages, supported by UNESCO. In the third edition of “Index for inclusion”, the authors Booth and Ainscow provide a set of materials to support the development of a formal school plan that reflects inclusive values. According to the authors “the Index can be integrated into this planning process by structuring a detailed review of the school and its relationship to its communities and environment, involving staff, governors, parents/carers and children” and “such a process itself contributes to the inclusive development of the school” (Booth & Ainscow, 2011). The authors outline three main dimensions when create a planning framework: cultures, policies and practices:

- policies are concerned with how the school is run and plans to change it;
- practices are about what is learnt and taught and how it is learnt and taught;
- cultures reflect relationships and deeply held values and beliefs. Changing cultures is essential in order to sustain development (Booth & Ainscow, 2011).

The authors divide each dimension into two sections, outlining the following planning framework (Booth & Ainscow, 2011):

- Dimension A: Creating inclusive cultures
  - A1: Building community
  - A2: Establishing inclusive values

- Dimension B: Producing inclusive policies
    - B1: Developing the school for all
    - B2: Organising support for diversity
- Dimension C: Evolving inclusive practices
    - C1: Constructing curricula for all
    - C2: Orchestrating learning (Booth & Ainscow, 2011).

For each section the “Index for inclusion” outlines a set of indicators, and for each indicator is developed a checklist of questions that define its meaning and provide a detailed review (Booth & Ainscow, 2011). The “Index” provides indicators that affect the process of inclusion at the school level in all aspects. The current study is focused on deriving criteria for assessing specifically the inclusive potential and inclusive power of learning scenarios. Therefore, for its purposes, mainly some of the indicators included in sections A1: “Building community”, A2: “Establishing inclusive values'' and C2: “Orchestrating learning” were extracted from the “Index”. The most important indicators related to the inclusive potential and inclusive power of learning scenarios in this aspect are “learning activities encourage the participation of all children”, “staff and parents/carers collaborate”, “children are valued equally”, “children help each other”, the school “encourages respect for all human rights”, “counters all forms of discrimination”, “promotes non-violent interactions and resolutions to disputes''

The identified, in the literature review, criteria for evaluation of inclusive potential were divided in several main groups, related respectively to the learning content, activities and instructions, assessment, participation and accessibility, equality and consideration of individual needs, cooperation and / or collaboration, social and personal relationships. The criteria also were related to the corresponding indicators defined in the “Index for inclusion”. Table 1 summarized the identified criteria and indicators by groups.

Table 1. Summary of the criteria for evaluation of inclusive potential by groups extracted from literature review

Groups of criteria	Criteria	Indicators
Content, instructions, assessment	Content and activities: meets the learning objectives/academic goals; activates prior experiences and knowledge relevant to the topics, strategies or skills to be learned; integrates lessons content or experiences; forecasts upcoming learning content; reference to the real world.	Children are encouraged to be confident critical thinkers; Lessons develop an understanding of the similarities and differences between people; Homework is set so that it contributes to every child’s learning; Staff develop shared resources to support learning; Resources in the locality of the school are known and used.

Groups of criteria	Criteria	Indicators
	Instructions: contains clear explanations; contains pre-set clear instructions, expected results and execution time; delivers instructional cues and prompts; provides error correction procedures; provides guidance on how to use; provides content-related assistance; avoid lengthy instructions; distinguishes clear individual stages of task performance; requires small steps of task performance	Teaching assistants support the learning and participation of all children.
	Assessment: summarizes accomplishments of individuals and group.	Assessments encourage the achievements of all children.
Participation	Contains elements that encourage the participation of all students; explicit requires the participation of all students in the class / group / team; involve parents support and/or participation; encourage parent involvement; provokes active involvement; provokes positive emotions.	Learning activities are planned with all children in mind; Learning activities encourage the participation of all children; Children are actively involved in their own learning; Staff and parents/carers collaborate; Staff and governors work well together; Staff link what happens in school to children’s lives at home.
Accessibility	Accessible to all students; provides engagement of all students according to their individual ability; allows to engage students in specific activities outside of the regular classroom; encourages membership of clubs, teams; encourages in and out of class activities; removes barriers and inequalities.	Activities outside school lessons involve all children.
Equality and consideration of individual needs	Applies the rules equally for all; support students according their individual needs; all learners are valued and respected.	Inclusion is viewed as increasing participation for all; Expectations are high for all children; Children are valued equally.
Cooperation and/or collaboration	Support cooperative/collaborative learning; uses appropriate incentives and rewards to encourage inclusion and collaboration; encourages collaboration of students to fix errors rather than providing directly the answer; involves working in groups to encourage cooperation between students; requires all students to work	Children learn from each other; Staff plan, teach and review together; Staff co-operate; Children help each other; All new children are helped to settle into the school.

Groups of criteria	Criteria	Indicators
	together (to complete the task); encourage collaboration and shared responsibility; provokes contacts/interactions; promotes social inclusion.	
Social and personal relationships	Focuses on the relationships between students; includes rules that require respect and support for other students; contains elements that encourage support between students; provokes peer support and acceptance; helps build friendships/relationships; helps to avoid bullying.	Discipline is based on mutual respect; Everyone is welcomed; Staff and children respect one another; The school is a model of democratic citizenship; The school encourages an understanding of the interconnections between people around the world; The school develops shared inclusive values; The school encourages respect for all human rights; The school encourages respect for the integrity of planet earth; The school counters all forms of discrimination; The school promotes non-violent interactions and resolutions to disputes; The school encourages children and adults to feel good about themselves.

5.2 Empirical research on PLEIADE designs refined groups; main criteria and sub-criteria

5.2.1 Pilot Learning Designs

During the PLEIADE project blended training activities (BTA), all teachers had to describe, design and assess inclusive learning scenarios. To do this, teachers, divided into six groups, had first to describe an inclusive and collaborative learning activity that took place in their schools (as shown on Table 2). The PLEIADE team proposed a template, conformed to the 4T model: where every collaborative learning activity is a task to be accomplished by one or more teams of learners within a certain time frame in a given technological environment (Persico & Pozzi, Task, Team and Time to structure online collaboration in learning environments, 2011), (Pozzi, Hofmann, Persico, Stegmann, & Fischer, 2013), (Pozzi & Persico, Sustaining learning design and pedagogical planning in CSCL, 2013). As the 4T model, developed by CNR-ITD, illustrates (Passarelli, Dagnino, Persico, Pozzi, & Nikolova, 2021) the collaborative learning scenario template covers the Task to be accomplished by the students and the production of a final output; the Teams where

students collaborate and how they interact; the Time to accomplish the task; and the Technology to carry out the activity and the interactions. Additionally, every learning scenario had to exploit one of the six collaborative techniques: jigsaw, roleplay, pyramid, peer review, discussion, and case study. Thus, PLEIADE participants produced 6 learning scenarios, as described in Table 2 (Passarelli, Dagnino, Persico, Pozzi, & Nikolova, 2021), (Pozzi, Ceregini, & Persico, 2016).

After the feedback session, teachers worked on several turns on their designs, in order to improve and to address all suggestions and comments.

Table 2. Description of Learning Designs (Passarelli, Dagnino, Persico, Pozzi, & Nikolova, 2021), (Pozzi, Ceregini, & Persico, 2016)

Design	Country	Description	Social Inclusion Issues	Collaborative technique
Interview master	BG	TV Interview about the noun - make a summary and revision of the grammar unit (noun)	Minorities: Overcome language barriers, Low self-esteem, Students from low social and economic background, pupils neglected by their parents	Roleplay
Desert Island	BG	A traveller who wound up on a desert island and has only one object with him to survive or escape from the island. Improve communication between students, verbal communication, storytelling, learning new words more easily, tolerance and patience to listen to all ideas.	Students who are shy and cannot express themselves, students with poor vocabulary	Peer Review Discussion Roleplay

Design	Country	Description	Social Inclusion Issues	Collaborative technique
My 1821 hero	GR	Learn about the life and achievements of a Greek Revolution (1821) Hero	Students with learning difficulties Socially distanced children that had lower participation and inclusion during the pandemic Overcome language barriers	Peer Review Discussion
Town Ghosts	IT	The activity aims at creating a treasure hunt in Augmented Reality with geotagging so that important people from the town would appear at relevant sights of the town to talk about the history of Trani.	Foreign students who took advantage of the use of English as vehicular language; Students with social and cultural disadvantage Special need students	Peer Review Jigsaw Discussion Case Study
Learning about the mass	CY	Understand the meaning of mass Learn how to use the measurement instrument Understand the unit of measurement of mass	Overcome language barriers Help to socialize refugee/migrant students to the other students	Roleplay Other
Pulmonary and systemic circulation	CY	Understand the workings of the double circulation in mammals Trace the pathway of the oxygenated and deoxygenated blood Engage all the students of the class	Overcome language barriers (Romanian-speaking and Arab-speaking students) Socialize refugee/migrant students	Roleplay

5.2.2 Extracting indicators for Inclusive potential from the first PLEIADE designs

The next step of the BTA activity consisted of a peer review among teachers’ groups (Pozzi, Ceregini, & Persico, 2016). Every group has to provide both quantitative and

qualitative evaluations of the others’ group learning designs. Teachers have to figure out their own criteria and reflections about inclusiveness and collaborative potential. Some guidance was provided by the PLEIADE team, focusing on the: 1) Clarity of description (understandable and reusable); 2) Coherence of the learning path/activity, 4T elements (time, task, team and technology) are well selected; 3) Inclusion issues; 4) Inclusion assets; 5) Collaborative techniques in the learning path/activity (collaborative techniques); 6) Suggestions for improving the activity.

Each school team, guided by CNR-ITD, provided its reasons and justification for the voting, and the analysis in Table 3 shows that the inclusive potential can cover specific issues.

Table 3. Summary of the Qualitative peer evaluation of the Learning Designs

Peer review	**Inclusive & Collaborative potential, values by peers and CNR-ITD experts**	**Suggestions for improvements, provided by peers and CNR-ITD experts**
Interview master	· **Roleplay** – each student had a role to play and contribute, fun and playful activity. · Calibrated to students with **special needs**. · Exportability of the design - new contexts, new contents. · No need for complex tools. · **All students** are involved to participate. · Mixed ability classes. **Advanced students help the weaker ones**.	· **Roles** are not clearly stated. · The activity is vague and has inadequate targets. · Little use of technology · The students towards whom the inclusion activity was aimed **were not actively involved**. More students should be implicated. · Suggestion: to make an **interchange of roles**
Desert Island	· A high level of inclusion – **all students have to discuss and negotiate the solution** to their problem. · Each student had to present his/her object of choice and **stand in front of the peers as equal**, feeling accepted as part of the team. · All students with difficulties or not have a chance to express their opinions. · The **interaction between the team members** before the actual	· Possible risk - **the most extroverted students can overshadow** the shier ones. · How **shy students** could "dare" to express themselves and be exposed to a broader audience? · Teachers have to provide more **strict criteria for peer assessment** in order to avoid any **biases by personal reasons** (for instance friendship). · There was no use of technology.

Peer review	Inclusive & Collaborative potential, values by peers and CNR-ITD experts	Suggestions for improvements, provided by peers and CNR-ITD experts
	presentation has the highest inclusive potential. · The activity uses **different codes**, from the linguistic to the graphic one and enhances communication skills (use drawing and arts instead of verbal communication).	· Not **clear team structure**: plenary, individual student or small groups.
My 1821 hero	· The activity is inclusive because **all students have their task** and each group is formed by children with **mixed abilities** and different features. · A school-wide project in a distance learning context (involving 130 students). · The jigsaw is a **collaborative strategy** which allows the spread of knowledge to all students. · The activities are good practice to be used also in one single class or with different subjects. · Students who are **extroverts can take part as presenters and students who are shy can work on the individual part of the activity**. The activity includes cooperation. · Technology: Teams, Microsoft SWAY-PPT, Microsoft Word, Microsoft Edge.	· Lack of teacher **supervision** about **how students are involved** and contribute to the task. · Inclusion risks: in case of lack of devices or Internet connection in students’ families, lack of basic digital skills. · Suggestions: · To have **4 students per group (instead of 5-6)**. · To provide to each group a list with **instructions** or specific problems to look for online. · To clarify how to **support** students that face various **obstacles** in their participation (learning disabilities, language barriers).
Town Ghosts	· Use of English language (a **foreign language for all students**), so they do not feel insecure regarding language matters since nobody uses their mother-tongue. · The activity can be applied for other purposes, in other subjects- science, literature, math, etc. It was playful, enjoyable and effective. · Students enhance their social skills since they **collaborate with other**	· It requires students to be really skilled regarding their age in carrying out all the tasks. · Although it was very well planned and easily applicable it lacked the necessary technology. · Note taking will be difficult for students **with learning difficulties** and it is not a task that promotes inclusion.

Peer review	Inclusive & Collaborative potential, values by peers and CNR-ITD experts	Suggestions for improvements, provided by peers and CNR-ITD experts
	students in a relaxed environment, work in teams and carry out research. · They participate both in the making but also in the execution of the game and **choose** the software. · The use of technology: Interactive Whiteboard using augmented reality, device connected to the Internet, Google Jamboard, VR.	· The vast majority of the schools (for ex. in Cyprus) don’t have this kind of technology. For this reason, this activity is not applicable for Cyprus Schools.

Learning about the mass	· The strategy of this design is very useful to foster inclusion;
	· **Roleplay** and making students use objects to learn about mass and scale, interacting as in **lifelike situations**, can be exploited in different contexts
	· **Enhances students' collaboration** through experimental techniques which entails a freer approach.
	· Use of **technology** (google translator) which **reinforces students’ self-esteem** since they also use their mother-tongue.
	· **Feedback** is provided. The **description is clear enough** and it can be understood.
		· The immigrant students could learn the words in different languages, but there weren't Greek students so it’s not a real inclusion activity.
		· The **time** spent on **individual work**
		· Except for language barriers, the activity is not very **collaborative**.
		· Too much **emphasis** on the translation of the words but not on the meaning and the definitions.
		· **Individual** work rather than in **groups**.
		· Suggestions:
		  · Introduce **role play**: “At the market”.
		  · Use sentences where students can use the terms in real world situations or in short dialogues.

Pulmonary and systemic circulation	· The activity allows students to overcome the linguistic barrier and to develop social skills (non-linguistic communication is prevalent). · **Roleplay: students were assigned "roles" such as heart, blood**. Pupils collaborate in **dramatising** their	· **Communication** is not addressed in the activity since the students **dramatise** the gas exchange in the human body without the use of dialogues. · The main resource is the possibility of dramatising a

Peer review	Inclusive & Collaborative potential, values by peers and CNR-ITD experts	Suggestions for improvements, provided by peers and CNR-ITD experts
	roles in the oxygenation cycle of the human body. · All students are included, they work in **groups** and **help each other**. · The presentation of the activities is clear and detailed with enough time previewed for the activities. · Students have the chance to deal with different tasks such as **drawing**, performing, using **symbols** and **colours** in order to express their point of view. · The foreign student has to verbalise the gas exchange and act as an “expert” for the group activity – filming is interesting and fun.	topic of biology using **non-verbal communication** without the need for knowledge of a common language. · Suggestions: · The PPT used to introduce the topic could be shortened to give space to the students to express themselves. · **Time** devoted to the activities seems to be short. · There was **inadequate use of technology**.

In addition to the peer-review, an expert evaluation of the designs was provided as well. The Pleiade team experts made both an in-depth revision of the text and commented on the scenario’s templates and additional comments, reflections, ideas and suggestions. Their comments allowed teachers to figure out their own designs in a more general context, to structure better the activities following the 4T model. The experts explored both the collaboration and the inclusive potential of each activity within the designs. Some of the key advice for the teachers included suggestions on how to make designs more universal and collaborative with more playful elements.

The feedback records not only help for improvement of the scenarios from inclusive aspect; they also provide a picture on which are the key **features** of the scenario that the scientists (i.e. CNR-ITD experts) and the practitioners (peers) look for. The intersection of both points of view presents a set of indicators, which are theoretically substantiated and at the same time intuitively felt by teachers based on their experience and practice, which shows that it will be easy and natural for them to use these indicators in the future to evaluate the inclusive potential (at the beginning) and inclusive power (later) of a learning design, they are working on.

A horizontal cut of the feedback – commented positive features and suggestions for improvement highlight several important factors determining the inclusive potential of the scenarios:

- **Collaborative activities**: The reviewers emphasis on the collaborative teaching and learning techniques and activities such as role play, jigsaw, etc. Even in traditional classroom activities, they are looking for possibilities and level of collaboration ensured by the learning design.
- **Classroom management:**
  - **Time management**: Time dedicated to each phase should be enough for each student to be able to perform the task, as well as the collaborative activities to be completed in an effective way. On the other hand, the learning time should be compressed well, not leaving an opportunity for students to lose the focus on the lesson or just to feel bored not having meaningful tasks.
  - **Group management**: Work in groups supports collaborative activities if the group management is scaffolded with attention at the very beginning. Most of the collaborative activities are appropriate for a work in small groups (2 - 4 pupils), while the same activities in a larger group became unmanageable. If the task requires individual work, the role of the teacher is to ensure that all students have an equal opportunity to participate and to present in the best way. In the plenary sessions plan, the reviewers track whether the design provides opportunity for each student to express himself or his team, increasing his own level of self-esteem and self-confidence.
- **Communication**: The reviewers emphasise on the use of different types of communication – verbal and non-verbal. Drama, pictures and symbols are identified as effective ways to support verbal communication in groups of students with linguistic differences or other learning difficulties.
- **Assessment**: The appropriate assessment techniques increase the student self-esteem and avoid the subjectiveness in the process. The peer-review is identified as a powerful technique, providing additional value in terms of collaboration, learning from each other, self-confidence increase, development of skills for reasoning and support for others. Strict criteria description and clear instructions for assessment are indicators for correct process avoiding biases and encouraging achievements of all students.
- **Use of technologies**: The teachers and experts treat the technologies as facilitators in different aspects – as tools and environment supporting collaborative activities (shared documents, presentations, etc.), as mediators of effective communication (whiteboards, VR, presentations, webinars, etc.) as well as tools for providing more concrete and visualised learning content (drawings, VR, models, etc.)

- **Values:** Providing feedback, the reviewers put special attention on cultivating students' value system by ensuring the supportive environment (fun, relax, support by teacher or peers), stimulating respect to others, providing opportunity for pupils to take ownership of knowledge and skills they develop (having choice, going into roles, taking responsibility). Working on real-life problem situations is perceived as motivating for students and answering the question "Why do we learn this?".

5.2.3 Mapping of literature review results and the empirical analysis results

To map the extracted indicators for assessment of a learning design to the main criteria for inclusiveness of the learning, the feedback was elaborated again, taking into account the role of the reviewers - mentor (expert) or peer (teacher).

The mentors’ comments are given as guidance for the refinement and rework of the learning scenario using the four key words. Their two main emphases are collaboration potential and inclusive potential. Their recommendations are separated according to the main criteria groups.

The other data is gathered from the peer-reviews (fragmented in 4T model as well). The teachers are more detail-oriented in their reviews. Their comments refer to particular parts of the designs, as the students' roles, possible problems and constraints. The teachers’ perspective observes the weaknesses but the strengths as well. The issues they raise need to be taken into consideration by building, self- and peer-evaluation of learning designs.

Given that the comments are based on ready-made designs, they cannot be considered as a single point of reference. Rather, they build a scheme that can serve as a basis for a criterion matrix in creating future designs and their evaluation.

Both groups provided constructive feedback. Analysis of all the comments are summarised on Table 4. It is noteworthy that some recommendations were made independently by experts and colleagues (highlighted in colour), suggesting that they should find a place in the list of inclusion assessment criteria.

Table 4. Summary of the issues spotted and addressed by mentors and peer-review comments

Criteria Groups	Mentors feedback	Teachers feedback
Content, instructions, assessment	· choose collaborative technique(s) · specify the addressed issue	· scaffolding the students activities by providing a list of instructions for the students to follow

Criteria Groups	Mentors feedback	Teachers feedback
	· clear the focus · point the activity aim (be not too generic) · simplify the instructions · add links between the activities · give a connection between the tasks · revise timing · build a sharable and comparable scenario	· keep in mind that the use of technology is constrained in some schools · define the role of the technology in use · build a scenario applicable for other purposes as well
	· break the scenario in multiple phases (one explicit task per phase) with clear 4T model for each of them · prepare a clear teacher’s introduction about the topic	
	· specify the expected artefacts · prepare an evaluation criteria on quantitative and qualitative work of the participants · set the role of the peer feedback	· how is measured that the students achieve the goal? · how to guarantee an unbiased peer feedback
Participation	· problem solution and performance have to be a group activity · stick to group work activities instead of individual tasks · each group member to be involved at each stage especially by the final presentation of the artefact · motivating	· revise the tasks to be group work rather than an individual one · groups with mixed abilities · playful · enjoyable · effective · dynamic · life-like situation to be involved
Accessibility		· skills level according to the age of the participants · technology in use (competency level)

Criteria Groups	Mentors feedback	Teachers feedback
		· right role based on the person’s abilities
Equality and consideration of individual needs		· give adequate roles of shy and introverts students · stimulate the extroverts without overshadowing the others · define how the students with disabilities or/and disadvantages take part and express themselves · assure that every member feel as a part of the group
Cooperation and/or collaboration	how different activities give a way for members to communicate effectively	· verbal and non-verbal communication · stimulate the imaginations of the participants · involve cognitive skills · aim a competency enhancement of each student
Cultivating a value system for development of social and personal relationships and attitude to collaboration		· feel important · feel accepted · boding · support each other · feel rewarded

5.3 Directions for use of the criteria for development and implementation for inclusive learning scenarios

Combining the literature review and the empirical study of the pilot scenarios and feedback, the following list of criteria and indicators for evaluating the inclusive potential and inclusive power was extracted (Table 5).

According to the methodology, the list is serving as a base for development of the initial questionnaire which had to be structured, refined, and operationalized through the Delphi study in order to produce simple, easy to use tools for evaluating the **inclusive potential** and **inclusive power** of the teaching designs.

Table 5. Initial criteria - a base for Delphi study

Groups of criteria	Criteria	Indicators
Content, instructions, assessment	**Content and activities:** • Meets the learning objectives/academic goals; • Activates prior experiences and knowledge relevant to the topics, strategies or skills to be learned; • Integrates lessons content or experiences; • Forecasts upcoming learning content; • Reference to the real world	• Children are encouraged to be confident critical thinkers; • Lessons develop an understanding of the similarities and differences between people; • Learning activities are focused on the learning goals and objectives • Different activities and tasks are naturally connected • Homework is set so that it contributes to every child’s learning; • Staff develop shared resources to support learning; • Resources in the locality of the school are known and used • The content is provided in concrete and visualised way. • Solving real life problems • Time, devoted to each activity, is appropriate for successful accomplishment of the task.
	• Instructions: • Clear explanations; • Pre-set clear instructions, expected results and execution time; • Delivers instructional cues and prompts; provides error correction procedures; • Provides guidance on how to use;	• Teaching assistants support the learning and participation of all children • The learning design is divided in multiple phases with clear 4T model for each of them • The expected artefacts are specified

Groups of criteria	Criteria	Indicators
	• Provides content-related assistance; • Avoid lengthy instructions; • Distinguishes clear individual stages of task performance; • Requires small steps of task performance	
	Assessment: • Summarizes accomplishments of individuals and group	• Assessments encourage the achievements of all children • Assessment methods contribute to increase the student self-esteem and avoid subjectivity in the process. • Assessment methods provide opportunity for measuring if the learning goals are achieved • Collaborative assessment techniques are used – peer review, discussions, feedback in teams, etc. • Strict assessment criteria avoiding biases
Participation	• Contains elements that encourage the participation of all students; • Explicit requires the participation of all students in the class / group / team; • Involve parents support and/or participation; • Encourage parent involvement; • Provokes active involvement; • Provokes positive emotions	• Learning activities are planned with all children in mind; • Learning activities encourage the participation of all children; • Learning activities in groups require active interaction between team members • Children are actively involved in their own learning; • Each group member is involved at each stage

Groups of criteria	Criteria	Indicators
Accessibility	• Accessible to all students; • Provides engagement of all students according to their individual ability; • Allows to engage students in specific activities outside of the regular classroom; • Encourages in and out of class activities; removes barriers and inequalities	• Activities outside school lessons involve all children • Usage of different codes – colours, pictures, drama, etc. • Effective verbal and non-verbal communication • Support for students facing various obstacles is ensured • Level of developed skills consider the age of participants • Extroverts and introverts are stimulated to express in relevant way • It is defined how the students with disabilities or/and disadvantages take part and express themselves • Every member feel a part of the group
Equality and consideration of individual needs	• Applies the rules equally for all; • Support students according their individual needs; • All learners are valued and respected	• Inclusion is viewed as increasing participation for all; • Expectations are high for all children; • Children are valued equally; • All students discuss and negotiate solutions together • Constructive feedback is provided to all students
Cooperation and/or collaboration	• Support cooperative/collaborative learning; • Uses appropriate incentives and rewards to encourage inclusion and collaboration; encourages collaboration of students to fix	• Collaboration learning strategies and techniques are planned / used • Children learn from each other; • Staff plan, teach and review together; • Staff co-operate;

Groups of criteria	Criteria	Indicators
	errors rather than providing directly the answer; ● Involves working in groups to encourage cooperation between students; ● Requires all students to work together (to complete the task); ● Encourage collaboration and shared responsibility; ● Provokes contacts/interactions; promotes social inclusion ● Ensures effective communication at all levels	● Children help each other; ● All new children are helped to settle into the school ● Active involvement of students at risk of exclusion ● Forming groups of students with mixed abilities ● Support of collaborative activities and communication by relevant ICT usage ● Aims a competence enhancement of each student ● Cooperation through verbal and non-verbal communication ● Staff and parents/carers collaborate
Classroom management	● Time management	● Time dedicated to each phase is enough for each student to be able to perform the task ● Properly defined time for collaborative activities to be completed in an effective way ● Well compressed learning time, not leaving an opportunity for students to lose the focus on the lesson or to feel bored.
	● Group management	● Appropriate form of work for each activity is chosen – individual, small groups, plenary to provide different levels of collaboration ● Work in groups supports collaborative activities ● In the plenary sessions, each student has opportunity to

Groups of criteria	Criteria	Indicators
		express himself or his team, increasing his own level of self-esteem and self-confidence
	• Environment management	• Learning with fun • Playful learning • Supportive environment • Technologies are used as tools and environment supporting collaborative activities (shared documents, presentations, etc.) • Technologies are used as a media for effective communication • Stimulates the imaginations of the participants
Cultivating a value system for development of social and personal relationships and attitude to collaboration	• Focuses on the relationships between students; • Includes rules that require respect and support for other student; • Contains elements that encourage support between students; • Provokes peer support and acceptance; • Helps build friendships/relationships; • Helps to avoid bullying	• Discipline is based on mutual respect; • Everyone is welcomed; • Staff and children respect one another; • The respect to others is stimulated • All students have opportunity to take ownership of knowledge and skills they develop by having a choice, taking responsibility • Learning content is connected to real life situations. • Each student feel important • Each student feel accepted • Each student feel rewarded • The understanding of interconnections between

Groups of criteria	Criteria	Indicators
		people around the world is encouraged ● Development of shared inclusive values; ● Respect for all human rights is encouraged; ● Respect for the integrity of planet earth is encouraged; ● Different forms of discrimination is taken into account ● It is promoted non-violent interactions and resolutions to disputes; ● Pupils are encouraged to feel good about themselves

6 Development of a structured list of criteria and sub-criteria to be incorporated in IO4 Evaluation kit

In this section, we describe the approach adopted to identify the list of criteria and sub-criteria that was incorporated in the IO4 evaluation kit in order to evaluate the inclusiveness of a collaborative learning design. The partnership decided to adopt the Delphi method to identify and contextually validate this list. Delphi (Landeta, 2006) is a consolidated research method based on a systematic procedure for building consensus on the solution of complex issues through consultation with a panel of experts. The Delphi method normally requires a number of rounds of consultation with the panel, through questionnaires constructed by the researchers with the aim of collecting the experts' opinion, while informing them at each round of the results of the previous consultation, so that individual members take into account the opinion expressed collectively in the previous round. An important feature of this method is reciprocal anonymity of the panel’s experts, which guarantees that the experts’ mutual influence is not biased by social norms (like assertiveness or fame of some members) but rather on deep reflection on peers’ opinions. In the Delphi method, the number of rounds is not decided beforehand, as the process ends when the new rounds do not lead to significant progress and consensus is reached. It should also be noted that the panel should stay the same throughout the whole process, with the only well known issue that some members may withdraw from the panel and therefore, at each round, the panel size is smaller. Thus, it is advisable that the panel members are formally committed to the task from the beginning and that their initial number is big enough to accommodate for withdrawals and still preserve an acceptable panel size.

Given the above mentioned features, the Delphi method is notoriously time consuming, due to the time needed by the researchers to analyze the data and prepare a new questionnaire for the following round and by the experts to reflect on the data emerged from the previous round and answer the new questionnaire.

Thus, it was decided that the kit development and the identification of the list of criteria and sub-criteria would run in parallel as much as possible, in order to produce the kit in the timeframe available for IO4 development. Hence, the UniSofia and CNR-ITD team worked in parallel according to the scheme of work represented in Fig. 2. While UniSofia carried out an initial literature review on criteria for developing inclusive designs (Zafirova-Malcheva, Antonova, Stamenkova, Nikolova, & Mihnev, 2022), the CNR-ITD team used the Nominal Group Technique to elicit from a subset of PLEIADE teachers participating in SJSTE3 their beliefs on inclusive pedagogies (Passarelli, Dagnino, Ivanov, & Persico, 2022). The results of these two

activities were merged into a first draft list of criteria and sub-criteria that was used as the starting point for the Delphi study. Thus, CNR-ITD run the first round of the Delphi and, based on its results, produced a second draft list to be used as an entry point for the second round of the Delphi. Thus, the UniSofia team developed a first prototype of the kit incorporating the second draft list of criteria and sub-criteria that was the outcome of the Delphi first round. This prototype was tested by the PLEIADE teachers during TPM6 in Athens. Meanwhile, the second Delphi round was run by CNR-ITD, and its outcome was a consolidated list of criteria and sub-criteria that was incorporated in the final prototype amended by the UniSofia team based on the feedback received in Athens. The third and last round of the Delphi was meant to collect the experts’ opinions about the weight of each criteria and sub-criteria. This Round, at the time of writing, is yet to be carried out, but it will not entail any change to the list, it will only provide information about the individual criteria’s weight. Thus, the finalisation of the kit was possible and the Delphi third round will be carried out beyond the completion of this Intellectual Output.

In the following, we will focus on the two rounds of the Delphi study that have been carried out so far, by providing details about the way they were carried out. The third and last round of the Delphi study is represented with dotted lines because it will likely be completed after the end of the project.

6.1 The PLEIADE Delphi study

6.1.1 Reasons for choosing the Delphi method

As mentioned above, the aim of the PLEIADE Delphi study is to identify a structured list of criteria and sub-criteria to evaluate a collaborative learning design in order to understand whether it can stimulate inclusive learning dynamics. The goal of the evaluation is not so much to summatively "measure" the inclusiveness of learning design, but rather to produce a diagnosis encouraging reflection on the strengths and weaknesses of the design, with regard to the inclusion of the PLEIADE students target, i.e. students who are socio-economically or culturally disadvantaged.

In line with the basic principles of the PLEIADE project (Persico, et al., 2023), the rationale for this study is based on two considerations. The first is that although collaborative learning is considered by many researchers as one of the most effective strategies in terms of inclusion (Floretta, 2021), (Herbert, 2011), (Rose, 2000), (Chita-Tegmark, 2012), its effectiveness in this regard cannot be taken for granted as it can be greatly amplified or hindered depending on the characteristics of the learning design and the way it is implemented (Pozzi, Manganello, & Persico, Collaborative Learning: a design challenge for teachers, 2023). The second

consideration is that the aim of the PLEIADE project, i.e. the development of teachers’ competence in the field of inclusive education, is best achieved through the intertwining of practice and reflection (Perla & Martini, 2019), (Borko, Jacobs, & Koellner, 2010), (Guskey, 2020), as well as practice sharing among teachers (Trust, Krutka, & Carpenter, 2016), (Lieberman & Pointer Mace, 2009), (Patton & Parker, 2017), (Persico, Milligan, & Littlejohn, The interplay between self-regulated professional learning and teachers’ work-practice, 2015), (Persico, Passarelli, Manganello, Gewerc Barujel, & Rodriguez Groba, 2023). As a consequence, the choice of the Delphi method, i.e. a participatory approach involving teachers and researchers as experts, can stimulate both participants in the Delphi study and the teachers who will use IO4 and the criteria it incorporates to reflect on their own inclusive practices and enrich them by leveraging the competence of other teachers and experts.

In line with these ideas, Fabbri, Striano and Melacarne (Fabbri, Striano, & Melacarne, 2008) suggest that effective teacher professional development should be based on devices that allow the retrieval of their teaching experience and the activation of a posteriori reflection on it.

Based on these considerations, the Delphi method was seen as a suitable method to develop the list of criteria because of its highly participatory approach. Thus, the list of criteria and IO4 itself are being developed by expert teachers and researchers with the aim of sharing with other teachers and researchers a device (IO4) that incorporates collective knowledge about inclusion and can activate reflection on it.

6.1.2 The Delphi method in PLEIADE

The PLEIADE Delphi study began in summer 2022 and is expected to be completed by the end of 2023. As mentioned above, in PLEIADE, agreement on a consolidated list of criteria was reached in two rounds, while the third round, at the time of writing, is yet to be carried out. This consolidated list is the one that has been incorporated in the IO4 kit. In the following, we provide details about the panel composition and then describe the two rounds in terms of input data, data collection tools and results.

6.1.2.1 Panel set up

The first step of a Delphi consists in setting up a reasonably big, but still manageable panel, with representatives of different types of expertise. In PLEIADE, the experts were recruited by invitation, either because they were known to some of the partners or because they were authors of relevant scientific publications in the field. Most of them also have significant teaching experience. Care was taken to include different kind of expertise relevant to PLEIADE’s aims, like intercultural

education, gender fairness, inclusion of Roma students, special needs education and gifted students’ inclusion. In total, the Delphi could rely on a panel consisting of 51 members (37 F, 14 M) from 14 different countries, mostly European who replied to the first round. Respondents to the second round amount to 39, two of which did not complete the questionnaire.

Not all the experts who joined the panel agreed with the publication of their names. Thus, the list of panel experts who agreed with the publication of their names will be made public through the project website once the third round of the Delphi study is completed, to respect the constraint of reciprocal anonymity of the experts.

6.1.2.2 First round method

As mentioned above, with the first round, the input data was a first draft of the list of criteria and sub-criteria derived from a literature review of the field (Zafirova-Malcheva, Antonova, Stamenkova, Nikolova, & Mihnev, 2022) and from a consultation with Greek, Bulgarian Cypriots, and Italian teachers participating in the PLEIADE project (Passarelli, Dagnino, Ivanov, & Persico, 2022) carried out using the Nominal Group Technique.

This list is reported in Table 6 and is made up of 5 criteria and their sub-criteria (totalling 42). The first 4 criteria refer, respectively, to the ability to promote active participation by all students, foster a positive atmosphere in the classroom, promote collaboration, open the classroom to the outside world. The last criterion refers to the internal consistency of the learning design, which is essential to assess the quality of a design regardless of its inclusiveness.

Aim of the first round was to find out whether the experts of the panel believed that all the criteria in the first draft were important enough to be included, that their formulation was clear, that the list was complete and, in case, what criteria or subcriteria were missing. Thus, the data collection tool used for the first round consultation with the panel of experts consisted in a questionnaire asking each panel member to give their opinions on the importance of each criteria and sub-criteria (on a scale from 1 to 5, with 1=not at all important at all and 5=extremely important) and providing them with the opportunity to comment through open-ended responses on the criteria and sub-criteria of the first draft and on its structure. The introduction to the questionnaire explained the whole Delphi procedure, asked members to commit to it and introduced them to the first draft of the list.

The questionnaire was developed through the Google Forms tool and it can be found in Appendix 1.

Table 6. The structured list of criteria in input to the Delphi study

Capacity to promote active participation by all	
1	active participation by all is encouraged/motivated
2	active participation by all is required
3	students' individual strengths are leveraged
4	students individual needs and weaknesses are taken care of
5	students are offered choices (personalization)
6	barriers and inequalities are addressed
7	gender bias and other stereotypes are avoided
8	technological choices do not hinder access
9	assessment for learning is preferred to assessment of learning
Capacity to promote a positive class atmosphere	
1	all students are valued and respected
2	students support each other
3	achievements of all students are encouraged
4	the class atmosphere is relaxed
5	bullying or fights are prevented/avoided
6	the class atmosphere is playful
7	students and teachers emotions are positive
8	minorities are not isolated; there are no tight cliques
9	the teaching and learning approaches are adequate to foster social bonds
Capacity to promote collaboration	
1	collaborative learning is the core approach of the design
2	collaboration is mostly aimed at producing shared artefacts
3	interdependence and shared responsibility are fostered and rewarded
4	collaboration is incentivated and rewarded
5	tasks are conceived to promote collaboration

6	team size and composition facilitate collaboration
7	technological choices facilitate collaboration
8	timing is adequate to allow collaboration
9	the collaborative activities' design favour new contacts and interactions
10	assessment strategies summarize the accomplishments of both individuals and group
Capacity to open up the class to the local and global context	
1	contacts with other cultures are encouraged
2	understanding of different points of view is encouraged
3	The activity involves people or places outside the classroom, a local level
4	The activity involves people or places outside the classroom, at international level
Coherence, completeness and clarity of the design	
1	aims are relevant for target needs
2	aims, learning objectives and contents are consistent with one another
3	assessment is aligned with learning objectives
4	the aims are in line with the curriculum
5	the design is clear
6	the teaching and learning approach is adequate to achieving the learning objectives
7	aims specifically target inclusion
8	the teaching and learning approach is adequate to achieving inclusion
9	the design is complete
10	contents promote inclusion (i.e. promote reflection about diversity, while encouraging reciprocal understanding)

6.1.2.3 First round data analysis and results

The analysis was done with descriptive statistics for importance and qualitative methods for open comments.

Table 6. The structured list of criteria in input to the Delphi studyTable 7 to Table 11 show the results obtained for each sub-criterion of the 5 main criteria.

Table 7. Results obtained for each sub-criterion of criterion “Capacity to promote active participation by all”

	Sub-criterion	N	Average	Standard Deviation	Standard error	Confidence interval (+/-)
1	active participation by all is encouraged/motivated	50	4,84	0,37	0,05	0,11
2	active participation by all is required	49	3,69	1,14	0,16	0,33
3	students' individual strengths are leveraged	48	4,42	0,92	0,13	0,27
4	students individual needs and weaknesses are taken care of	48	4,6	0,82	0,12	0,24
5	students are offered choices (personalization)	50	4,42	0,84	0,12	0,24
6	barriers and inequalities are addressed	48	4,60	0,84	0,12	0,25
7	gender bias and other stereotypes are avoided	49	4,45	0,82	0,12	0,23
8	technological choices do not hinder access	49	4,51	0,68	0,1	0,2
9	assessment for learning is preferred to assessment of learning	46	4,28	0,83	0,12	0,25

Table 8. Results obtained for each sub-criterion of criterion “Capacity to promote a positive class atmosphere”

	Sub-criterion	N	Average	Standard Deviation	Standard error	Confidence interval (+/-)
1	all students are valued and respected	49	4,92	0,28	0,04	0,08
2	students support each other	50	4,46	0,86	0,12	0,25
3	achievements of all students are encouraged	49	4,61	0,7	0,1	0,2
4	the class atmosphere is relaxed	50	4,46	0,73	0,1	0,21

	Sub-criterion	N	Average	Standard Deviation	Standard error	Confidence interval (+/-)
5	bullying or fights are prevented/avoided	50	4,56	0,84	0,12	0,24
6	the class atmosphere is playful	50	3,88	0,98	0,14	0,28
7	students and teachers emotions are positive	49	4,39	0,79	0,11	0,23
8	minorities are not isolated; there are no tight cliques	50	4,72	0,57	0,08	0,16
9	the teaching and learning approaches are adequate to foster social bonds	48	4,52	0,77	0,11	0,22

Table 9. Results obtained for each sub-criterion of criterion “Capacity to promote collaboration”

	Sub-criterion	N	Average	Standard Deviation	Standard error	Confidence interval (+/-)
1	Collaborative learning is the core approach of the design	48	4,35	0.84	0.12	0.24
2	Collaboration is mostly aimed at producing shared artefacts	50	3,62	1.24	0.18	0.35
3	Interdependence and shared responsibility are fostered and rewarded	49	4,39	0.89	0.13	0.25
4	Collaboration is incentivated and rewarded	48	4,21	0.92	0.13	0.27
5	Tasks are conceived to promote collaboration	48	4,4	0.82	0.12	0.24
6	Team size and composition facilitate collaboration	49	4,35	0.90	0.13	0.26
7	Technological choices facilitate collaboration	49	4,29	0.87	0.12	0.25
8	Timing is adequate to allow collaboration	49	4,51	0.74	0.11	0.21
9	The collaborative activities' design favour new contacts and interactions	50	4,46	0.68	0.10	0.19
10	Assessment strategies summarise the accomplishments of both individuals and group	49	4,45	0.87	0.12	0.25

Table 10. Results obtained for each sub-criterion of criterion “Capacity to open up the class to the local and global context”

	Sub-criterion	N	Average	Standard Deviation	Standard error	Confidence interval (+/-)
1	Contacts with other cultures are encouraged	49	4,57	0,58	0,08	0,17
2	Understanding of different points of view is encouraged	50	4,82	0,44	0,06	0,12
3	The activity involves people or places outside the classroom, at local level	50	4,12	0,94	0,13	0,27
4	The activity involves people or places outside the classroom, at international level	50	3,78	0,95	0,13	0,27

Table 11.Results obtained for each sub-criterion of criterion “Coherence, completeness and clarity of the design”

	Sub-criterion	N	Average	Standard Deviation	Standard error	Confidence interval (+/-)
1	aims are relevant for target needs	48	4,625	0,70	0,10	0,20
2	aims, learning objectives and contents are consistent with one another	50	4,66	0,72	0,10	0,20
3	assessment is aligned with learning objectives	50	4,58	0,73	0,10	0,21
4	the aims are in line with the curriculum	47	4,09	0,95	0,14	0,28
5	the design is clear	49	4,69	0,58	0,08	0,17
6	the teaching and learning approach is adequate to achieving the learning objectives	48	4,08	1,01	0,15	0,29
7	aims specifically target inclusion	50	4,54	0,79	0,11	0,22
8	the teaching and learning approach is adequate to achieving inclusion	50	4,36	0,90	0,13	0,26

	Sub-criterion	N	Average	Standard Deviation	Standard error	Confidence interval (+/-)
9	the design is complete	48	4,71	0,58	0,08	0,17
10	contents promote inclusion (i.e. promote reflection about diversity, while encouraging reciprocal understanding)	47	4,49	0,80	0,12	0,24

According to the above data, the importance ascribed by the experts to each sub-criterion in Round 1 was very high on average, with the mean never below 3.6 on a scale of 1-*not at all important* to 5-*extremely important*. However, the analysis of the open-ended responses to the questionnaire revealed a number of misunderstandings as well as explicit suggestions for reformulation or addition of criteria and sub-criteria to be submitted to the experts in the next Round (concerning a total of about 80 percent of the starting criteria).

The open-ended answers were qualitatively analysed by four independent coders with a unit of analysis each individual suggestion provided, which often coincided with the answer. When there was agreement of at least two respondents on a proposed change, the researchers transformed them into “proposals'' to be submitted to the panel in the second round. The same happened when a suggestion came from only one respondent but at least two coders agreed that the suggestion had to be acted upon. The result of the analysis led to the formulation of a very extensive criteria revision proposal: most importantly, one criterion was proposed to be added to the initial 5 criteria. This additional criterion, “Contextual features and teachers' competences”, relates to the characteristics of the school context and the teacher's ability to manage the classroom in an inclusive manner. Its sub-criteria (see Table 12), although not inferable from the examination of the learning design, certainly have an important weight in determining the effectiveness of the teaching intervention with regard to inclusion. In addition to this major change, 18 sub-criteria rewordings were proposed, along with 15 new sub-criteria for addition, one for deletion, and two cases of merge.

Table 12. The newly proposed criterion emerged from round 1 and its sub-criteria

Contextual features and teachers' competences	
1	Family involvement is promoted by the teacher and the school.
2	The physical classroom setting facilitates interaction among students.
3	The teacher is skilled in classroom management.

4	The teacher frequently stimulates dialogue among students.
5	The teacher encourages active listening.
6	The teacher is attentive to the social and cultural values that may be implicitly conveyed (ideals, stereotypes, judgements and prejudices).
7	The teacher pays attention to non-verbal communication to identify inclusion issues.
8	The teacher's feedback is formative and constructive rather than summative and judgemental.

6.1.2.4 Second Round method

In the second round the 51 respondents of the first round were requested to express their opinion about those criteria and sub-criteria that, based on the previous round, were proposed for elimination, rewording, merge or addition.

Specifically, for each of the 6 criteria (including the newly proposed one), any change concerning them and their sub-criteria were submitted to the experts for judgement in Round 2. This round, the experts were asked to confirm or not confirm the changes proposed by some of their peers in the previous Round. The questionnaire was implemented in LimeSurvey and can be found in Appendix 2.

6.1.2.5 Second Round analysis and results

The analysis of the data emerged from round 2 consisted in approving only the changes agreed upon by the simple majority of the respondents. Thus, at the end of Round 2, one of the original sub-criteria was discarded (sub-criterion 2 of the first criterion, with 33 votes in favour against 6), most of the proposals for rewording some of the criteria were approved and the same happened with the proposals for merging and adding sub-criteria. The additional criterion in Table 12 was also confirmed, with all of its sub-criteria. The final version approved by the experts is as shown in Table 13 and includes 6 criteria and 53 sub-criteria that are now consolidated and have been incorporated in IO4.

Thus, the total number of criteria increased from 5 to 6 and that of sub-criteria from 42 to 53.

Table 13. The results of Delphi Round 2 and the structured list of consolidated criteria

Capacity to promote active participation by all		
	Results of Round 2	**Resulting sub-criteria**
1	New formulation preferred to old one (22 vs 18)	Active participation by all is encouraged/motivated through a student centered approach (e.g. game based learning).

2	New formulation preferred to old one (39 vs 1)	Active participation by all is encouraged/motivated through a student centered approach (e.g. game based learning).
3	Unchanged	Students' individual needs and weaknesses are taken care of.
4	New formulation preferred to old one (34 vs 6)	Students are offered choices in terms of media, content, learning strategies, or types of artefacts to produce (personalization).
5	Addition (34 vs 6)	Students are provided with opportunities to negotiate some design decisions (e.g., deadlines, assessment modes, assessment rubrics, learning objectives).
6	Unchanged	Barriers and inequalities are addressed.
7	Unchanged	Gender bias and other stereotypes are avoided.
8	New formulation preferred to old one (30 vs 10)	Materials and technology employed do not hinder access by the students involved.
9	New formulation preferred to old one (31 vs 9)	Assessment for learning (i.e. formative assessment) is preferred to assessment of learning (i.e. summative assessment).
10	New formulation preferred to old one (38 vs 2)	Students are actively involved in the assessment process (e.g., through peer feedback).
11	New formulation preferred to old one (37 vs 3)	Students are offered opportunities for personal expression.
Capacity to promote an open class atmosphere (reworded)		
	Results of Round 2	**Resulting sub-criteria**
1	Unchanged	All students are valued and respected.
2	Unchanged	Students support each other.
3	Unchanged	All students receive encouragement for their achievements.
4	Merge of previous sub-criteria 4 and 5 (...)	The class atmosphere is positive (e.g. relaxed, playful).
5	New formulation preferred to old one (37 vs 2)	Bullying is hindered and conflict management supported.
6	New formulation preferred to old one (37 vs 2)	Positive emotions are promoted while negative ones are recognised and managed.
7	Unchanged	Minorities are not isolated; there are no tight cliques.
8	Unchanged	Teaching and learning approaches are suitable for fostering social bonds.
9	Added (36 vs 3)	Mistakes are not stigmatized or punished, but used to enhance learning.
Capacity to promote collaboration		

	Results of Round 2	**Resulting sub-criteria**
1	Unchanged	Collaborative learning is the core approach of the design.
2	New formulation preferred to old one or delete (21 vs 3/15)	Collaboration is mostly aimed at producing shared artefacts (e.g. a performance, a presentation, a tangible object, a report).
3	Unchanged	Interdependence and shared responsibility are fostered and rewarded.
4	Unchanged	Collaboration is promoted and rewarded.
5	Unchanged	Tasks are designed to promote collaboration.
6	Unchanged	Team size and composition facilitate collaboration.
7	Unchanged	Adopted technologies facilitate collaboration.
8	New formulation preferred to old one (32 vs 7)	Timing is adequate and flexible enough to allow collaboration.
9	New formulation preferred to old one (28 vs 11)	The collaborative activities' design facilitate new contacts and interactions (e.g. through rotating teams).
10	Unchanged	Assessment strategies address the accomplishments of both individuals and groups.
11	Added (38 vs 1)	Assessment strategies take both the process and the product into consideration.
12	Added (37 vs 2)	Students are offered opportunities to assume and play different roles.
Capacity to open up the class to the local and global context		
	Results of Round 2	**Resulting sub-criteria**
1	New formulation preferred to old one (25 vs 14)	Contacts with other cultures are encouraged.
2	Unchanged	Understanding of different points of view is encouraged.
3	New formulation preferred to old one or delete (28 vs 7/4)	The activity involves people or places outside the classroom, at local/national level (face to face or virtually).
4	New formulation preferred to old one or delete (33 vs 3/3)	The activity involves people or places outside the classroom, at international level (face to face or virtually).
5	Added (26 vs 13)	Content written or designed by people from the cultures involved is provided.
General design features		
	Results of Round 2	**Resulting sub-criteria**
1	Unchanged	The aims are relevant to the students' needs.

2	Unchanged	Aims, learning objectives and contents are mutually consistent.
3	Unchanged	Assessment is aligned with the learning objectives.
4	Merge of sub-criteria 5 and 9 preferred to delete or keep original (24 vs 10/5)	The design is clear and complete enough for other teachers to reuse.
5	New formulation preferred to old one (25 vs 14)	The teaching and learning approach is appropriate for achieving the aims and learning objectives.
6	New formulation preferred to old one (28 vs 11)	Inclusion is explicitly mentioned among the aims.
7	New formulation preferred to old one (28 vs 11)	The teaching and learning approach is appropriate for achieving inclusion.
8	New formulation preferred to old one (33 vs 6)	Contents promote inclusion (i.e. promote reflection about diversity and inequities, while encouraging reciprocal understanding).
Contextual features and teachers' competences		
	Results of Round 2	**Resulting sub-criteria**
1	Added (28 vs 6)	Family involvement is promoted by the teacher and the school.
2	Added (31 vs 3)	The physical classroom setting facilitates interaction among students.
3	Added (32 vs 0)	The teacher is skilled in classroom management.
4	Added (24 vs 5)	The teacher frequently stimulates dialogue among students.
5	Added (36 vs 1)	The teacher encourages active listening.
6	Added (34 vs 1)	The teacher is attentive to the social and cultural values that may be implicitly conveyed (ideals, stereotypes, judgements and prejudices).
7	Added (36 vs 1)	The teacher pays attention to non-verbal communication to identify inclusion issues.
8	Added (26 vs 0)	The teacher's feedback is formative and constructive rather than summative and judgemental.

6.2 From the list of criteria to the IO4 kit

The PLEIADE Delphi study produced a structured list of criteria and sub-criteria by supporting the panel of experts in reaching a final version agreed upon by most experts. In the context of PLEIADE, these criteria are incorporated in IO4, a toolkit used to evaluate the inclusiveness of the learning designs produced by the teachers.

Most of the criteria identified are formulated in such a way that the evaluation can be carried out both before and/or after the learning design is put into practice in the

classroom. In the terminology of the PLEIADE project, we refer to the first type of evaluation using the term "inclusive potential," as its evaluation is based exclusively on the characteristics of the design. On the other hand, we refer to the second type of evaluation by using the expression "inclusive power” of the design, which is based on ex-post observation of the class dynamics.

In view of the development of a toolkit capable of supporting both the evaluation of the inclusive potential and that of the inclusive power, Table 14 has been produced, where for each criterion two questions are formulated: one addressing the former and one addressing the latter. Whenever one of the two questions is not present in Table 14, this means that the sub-criterion only lends itself for assessing one of the two. In particular, it should be noted that the sub-criteria of the newly added criterion “Contextual features and teachers' competences” are all suitable only for assessing the power.

Table 14. Assessment kit questions derived from the sub-criteria of each criterion

	Criterion1: Capacity to promote active participation by all		
	Sub-criterion	**Question for inclusive potential**	**Question for inclusive power**
1	Active participation by all is encouraged/motivated through a student centered approach (e.g. game based learning).	Is active participation by all students encouraged/motivated through a student centered approach (e.g. game based learning)?	Was active participation by all encouraged/motivated through a student centered approach (e.g. game based learning)?
2	Students' individual strengths and experiences are leveraged and promoted.	Are students' individual strengths and experiences leveraged and promoted?	Were students' individual strengths and experiences leveraged and promoted?
3	Students' individual needs and weaknesses are taken care of.	Are individual needs and weaknesses taken care of?	Were students' individual needs and weaknesses taken care of?
4	Students are offered choices in terms of media, content, learning strategies, or types of artefacts to produce (personalization).	Are students offered the possibility to choose media, content, learning strategies, or types of artefacts to produce (personalization)?	Did students choose media, content, learning strategies, or types of artefacts to produce (personalization)?
5	Students are provided with opportunities to negotiate some design decisions (e.g., deadlines, assessment modes, assessment rubrics, learning objectives).	Are students provided with opportunities to negotiate some design decisions (e.g., deadlines, assessment modes, assessment rubrics, learning objectives)?	Did students negotiate some design decisions (e.g., deadlines, assessment modes, assessment rubrics, learning objectives)?

6	Barriers and inequalities are addressed.	Are barriers and inequalities addressed?	Were barriers and inequalities successfully addressed?
7	Gender bias and other stereotypes are avoided.	Does the design help avoid gender bias and other stereotypes?	Did the design succeed in avoiding gender bias or other stereotypes?
8	Materials and technology employed do not hinder access by the students involved.	Are the materials and technology employed accessible for all students involved?	Were materials and technology employed accessible by involved students?
9	Assessment for learning (i.e. formative assessment) is preferred to assessment of learning (i.e. summative assessment).	Are forms of assessment for learning (i.e. formative assessment) preferred to forms of assessment of learning (i.e. summative assessment)?	Did the teacher prefer forms of assessment for learning (i.e. formative assessment) over forms of assessment of learning (i.e. summative assessment)?
10	Students are actively involved in the assessment process (e.g., through peer feedback).	Are students actively involved in the assessment process (e.g., through peer feedback)?	Were students involved in the assessment process (e.g., through peer feedback)?
11	Students are offered opportunities for personal expression.	Does the design offer students opportunities for personal expression?	Were students offered opportunities for personal expression?
	Criterion2: Capacity to promote an open class atmosphere		
	Sub-criterion	**Question for inclusive potential**	**Question for inclusive power**
1	All students are valued and respected.	Does the design ensure that all students are valued and respected?	Were all students valued and respected?
2	Students support each other.	Does the design encourage students to support each other?	Did students support each other?
3	All students receive encouragement for their achievements.	Does the design encourage all students to make achievements?	Were the achievements of all students acknowledged or rewarded?
4	The class atmosphere is positive (e.g. relaxed, playful).	Does the design promote a positive atmosphere (e.g. relaxed, playful)?	Was the class atmosphere positive (e.g. relaxed, playful)?
5	Bullying is hindered and conflict management supported.	Does the design hinder bullying and support conflict management?	Did the design succeed in hindering bullying and support conflict management?
6	Positive emotions are promoted while negative	Does the design support promotion of positive emotions and	Were positive emotions promoted and negative

	ones are recognised and managed.	identification and management of negative ones?	ones recognised and managed?
7	Minorities are not isolated; there are no tight cliques.	Does the design help to avoid isolation of minorities and consolidation of cliques?	Was isolation of minorities and consolidation of cliques successfully avoided?
8	Teaching and learning approaches are suitable for fostering social bonds.	Are the teaching and learning approaches suitable to foster social bonds?	Did the teaching and learning approach foster social bonds ?
9	Mistakes are not stigmatized or punished, but used to enhance learning.	Does the design help to avoid stigmatization of mistakes, but rather foster their use as opportunities to enhance learning?	Were mistakes used as opportunities for learning, instead of being stigmatized or punished?
	Criterion3: Capacity to promote collaboration		
	Sub-criterion	**Question for inclusive potential**	**Question for inclusive power**
1	Collaborative learning is the core approach of the design.	Is collaborative learning the core approach of the design?	Did the students spend a significant part of the enactment in collaborative learning activities?
2	Collaboration is mostly aimed at producing shared artefacts (e.g. a performance, a presentation, a tangible object, a report).	Is collaboration aimed at producing shared artefacts (e.g. a performance, a presentation, a tangible object, a report)?	Was collaboration aimed at producing shared artefacts (e.g. a performance, a presentation, a tangible object, a report)?
3	Interdependence and shared responsibility are fostered and rewarded.	Does the design foster and reward interdependence and shared responsibility?	Was there a high degree of interdependence and shared responsibility?
4	Collaboration is promoted and rewarded.	Does the design promote and reward collaboration?	Was collaboration promoted and rewarded?
5	Tasks are designed to promote collaboration.	Are the chosen tasks suited to promote collaboration?	Were the tasks effective in promoting collaboration?
6	Team size and composition facilitate collaboration.	Do team size and composition facilitate collaboration?	Did team size and composition facilitate collaboration?
7	Adopted technologies facilitate collaboration.	Are technological choices suitable to foster collaboration?	Did the chosen technology help to foster collaboration?
8	Timing is adequate and flexible enough to allow collaboration.	Is timing adequate and flexible enough to allow collaboration?	Was timing adequate and flexible enough to allow collaboration?

9	The collaborative activities' design facilitate new contacts and interactions (e.g. through rotating teams).	Does the design of collaborative activities facilitate new contacts/interactions (e.g. through rotating teams)?	Did the collaborative activities facilitate new contacts/interactions (e.g. through rotating groups)?
10	Assessment strategies address the accomplishments of both individuals and groups.	Do assessment strategies address the accomplishments of both individuals and groups?	Did assessment strategies address the accomplishments of both individuals and groups?
11	Assessment strategies take both the process and the product into consideration.	Do assessment strategies take both process and product into conderation?	Did assessment strategies succeed in taking both process and product into consideration?
12	Students are offered opportunities to assume and play different roles.	Are students offered opportunities to assume and play different roles?	Were students offered opportunities to assume and play different roles?
	Criterion4: Capacity to open up the class to the local and global context		
	Sub-criterion	**Question for inclusive potential**	**Question for inclusive power**
1	Contacts with other cultures are encouraged.	Does the design encourage contact with other cultures?	Were contacts with other cultures encouraged?
2	Understanding of different points of view is encouraged.	Does the design encourage understanding of different points of view?	Did students gain understanding of different points of view?
3	The activity involves people or places outside the classroom, at local/national level (face to face or virtually).	Does the design involve people or places outside the classroom, at local/national level (face to face or virtually)?	Did students meet people or visit places outside the classroom, at local/national level (face to face or virtually)?
4	The activity involves people or places outside the classroom, at international level (face to face or virtually).	Does the design involve people or places outside the classroom, at international level (face to face or virtually)?	Did students meet people or visit places outside the classroom, at international level (face to face or virtually)?
5	Content written or designed by people from the cultures involved is provided.	Does the design provide content written or designed by people from the cultures involved?	
	Criterion5: General design features		
	Sub-criterion	**Question for inclusive potential**	**Question for inclusive power**

1	The aims are relevant to the students' needs.	Are the design aims relevant to the students' needs?	Is there any evidence that the aims matched students' needs?
2	Aims, learning objectives and contents are mutually consistent.	Are the aims, learning objectives and content of the design mutually consistent?	
3	Assessment is aligned with the learning objectives.	Is assessment in line with the learning objectives?	Did assessment succeed in measuring achievement of learning objectives?
4	The design is clear and complete enough for other teachers to reuse.	Is the design clear and complete enough for other teachers to reuse?	
5	The teaching and learning approach is appropriate for achieving the aims and learning objectives.	Is the teaching and learning approach appropriate for achieving the aims and learning objectives?	Did the students achieve the aims and learning objectives?
6	Inclusion is explicitly mentioned among the aims.	Is inclusion explicitly mentioned among the aims?	Was the class more inclusive after the enactment?
7	The teaching and learning approach is appropriate for achieving inclusion.	Is the teaching and learning approach appropriate for achieving inclusion?	Did the teaching and learning approach prove appropriate for achieving inclusion?
8	Contents promote inclusion (i.e. promote reflection about diversity and inequities, while encouraging reciprocal understanding).	Do contents promote inclusion (i.e. promote reflection about diversity and inequities, while encouraging reciprocal understanding)?	Did the contents promote inclusion (i.e. promoting reflection about inequities and/or diversity, while encouraging reciprocal understanding)?
	Criterion6: Contextual features and teachers' competences		
	Sub-criterion	**Question for inclusive potential**	**Question for inclusive power**
1	Family involvement is promoted by the teacher and the school.		Did the teacher and school promote family involvement?
2	The physical classroom setting facilitates interaction among students.		Did the physical setting of the classroom facilitate interaction among the students?
3	The teacher is skilled in classroom management.		Was the classroom effectively managed?

4	The teacher frequently stimulates dialogue among students.		Did the teacher frequently stimulate dialogue among the students?
5	The teacher encourages active listening.		Did students listen actively?
6	The teacher is attentive to the social and cultural values that may be implicitly conveyed (ideals, stereotypes, judgements and prejudices).		Was the teacher attentive to social and cultural values implicitly conveyed while teaching (ideals, stereotypes, judgements and prejudices)?
7	The teacher pays attention to non-verbal communication to identify inclusion issues.		Was the teacher attentive to non-verbal communication for identifying inclusion issues?
8	The teacher's feedback is formative and constructive rather than summative and judgemental.		Was teacher feedback formative and constructive rather than summative and judgemental?

7 Evaluation kit development

The list of criteria and sub-criteria, produced as a result of Round 2 of the Delphi study, provided a grounded set of consolidated indicators, operationalizing assessment of a design (in terms of inclusive potential) and a teaching intervention (in terms of inclusive power). Once the analysis of Round 2 results was completed, the list of indicators were incorporated into the prototype kit.

As required by the proposal, the Evaluation kit should consist of a text file and an interactive resource. The type of interactive resources was discussed among the leaders UniSofia and CNR-ITD. As it should serve mainly teachers and teachers’ trainers, an easy to use, popular tool for implementation of the operationalizing lists was needed. The natural e-presentation of a list is an electronic sheet. In addition, they provide functionality to calculate data and to present it in a graphical w

Among different software solutions (MS Forms, Survey Monkey, Libre Office, etc.), the Google sheets were chosen. The main reasons are that it is free and popular. In addition, the Google sheets provides an easy way to generate Google Form from a sheet – a feature that provides the possibility to easily change the representation in case of need after collecting users’ feedback. The solution meet the proposal requirements – the kit should be *easily manageable, and widely shareable in the community of student teachers.*

The kit should also be *easily understood and clear in meaning*. The Delphi study ensures clear understanding of the criteria and sub-criteria, operationalized in terms of questions related to the

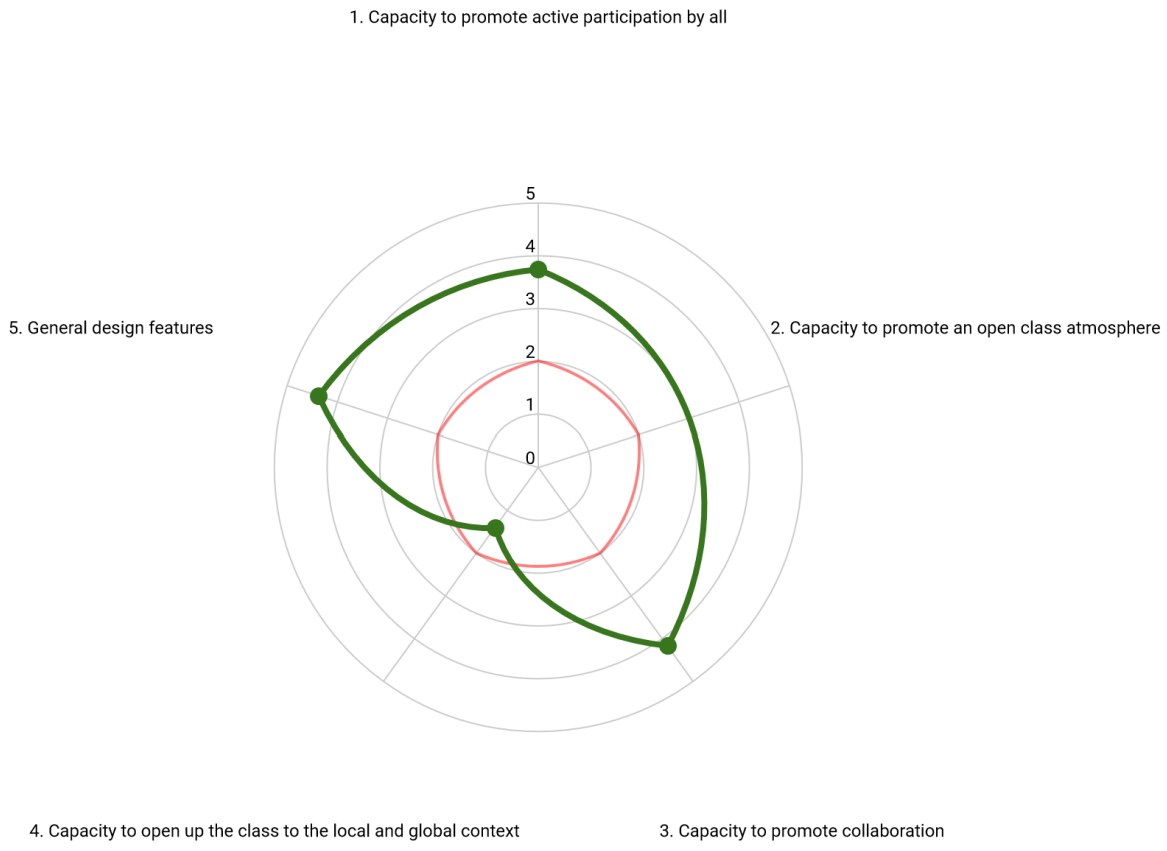


Figure . Radar char, presenting visual feedback as an integral part of the Evaluation toolkit

evaluation of the inclusive potential of a collaborative teaching design and of inclusive power of teaching via collaborative methods. To provide also a clear and easy to understand feedback by the kit, the summary of the results is presented in a graphical way, by a radar chart (Figure 3). Each ray of the radar represents a summary of a given criteria. A red line outlines the lowest acceptable border for each inclusive criteria. The green line represents the user's results of the evaluation. If the green line intersect the red one, there is at least one weak criterion. The visualization allows users to focus attention on particular weak criteria and to read again carefully the operationalized sub-criteria in order to find a solution for improvement of the design and / or enactment.

In addition, textual feedback in the form of suggestions, also helps the user to focus their attention on the weak criteria.

Having in mind the different context of usage of the evaluation kit, and the feedback taken during the validation process, two versions of the kit were developed – **Evaluation kit for individual user** and **Evaluation kit for multiple users**. Each kit contains two tools – for assessment of the **inclusive potential** of a collaborative learning design, and for assessment of the **inclusive power** of an enactment.

In this version all the sub-criteria are equally weighted. After the Round 3 of the Delphi study, they will be mapped to respective weights, reflecting on the experts' evaluation of importance.

7.1 Requirements to use the tool:

To use the tool, the user should be equipped with:

- Internet connection
- Web browser (such as Google Chrome, Microsoft Edge, Safari, Mozilla Firefox etc.)
- Google account to access Google Sheets

The Google account for individuals is free and can be created at https://accounts.google.com/.

7.2 Evaluation kit for individual user

7.2.1 Purpose

The purpose of the **Evaluation kit** for individual user is to operationalize the assessment of a design (in terms of **inclusive potential**) and a teaching intervention (in terms of **inclusive power**). The **Evaluation kit for individual user** supports individual experts (teachers, teachers' trainers, pre-service and in-service teachers, observers and other specialists) in many different contexts, such as:

- Evaluation of **inclusive potential** of a learning design in advance:
    - by a teacher, for self-assessment and improvement of a learning design (produced with the I4T game or otherwise);

  - by a teacher, for selection of a relevant design for enactment from the PLEIADE Good Practices Repository;
  - by a teachers’ trainer, for assessment of a student’s design;
  - by a pre- or in-service teacher, as a rubric for assessment of a peer’s design during the teachers’ training in a peer review process;
  - by a school manager / head teacher, for assessment of an young teacher design;
  - by a regional expert before observation of a teacher’s lesson.
- Evaluation of the **inclusive power** of teaching/learning process, achieved by the teacher’s intervention;
  - by a teacher, for self-reflection and improvement of the teaching practice;
  - by a teachers’ trainer during observation of the teaching practice of a student;
  - by peer, for peer review of the teaching practice of a colleague during teacher’s training;
  - by a regional or school manager during observation of a teacher’s practice

7.2.2 Access of the Evaluation kit for individual user

The Evaluation kit for individual user in English is accessible through the link: https://docs.google.com/spreadsheets/d/1e0xJcl6ObUqU1RdrL-6kLmDo7eNRA3kncFzGXJZwF_g/copy or through the QR code:

The link asks the user to create a copy of the original kit, so to have a private access to his own evaluation data.

7.2.3 Description of the Evaluation kit for individual user

The first sheet of the kit provides detailed instructions on how to use it (Figure 4).

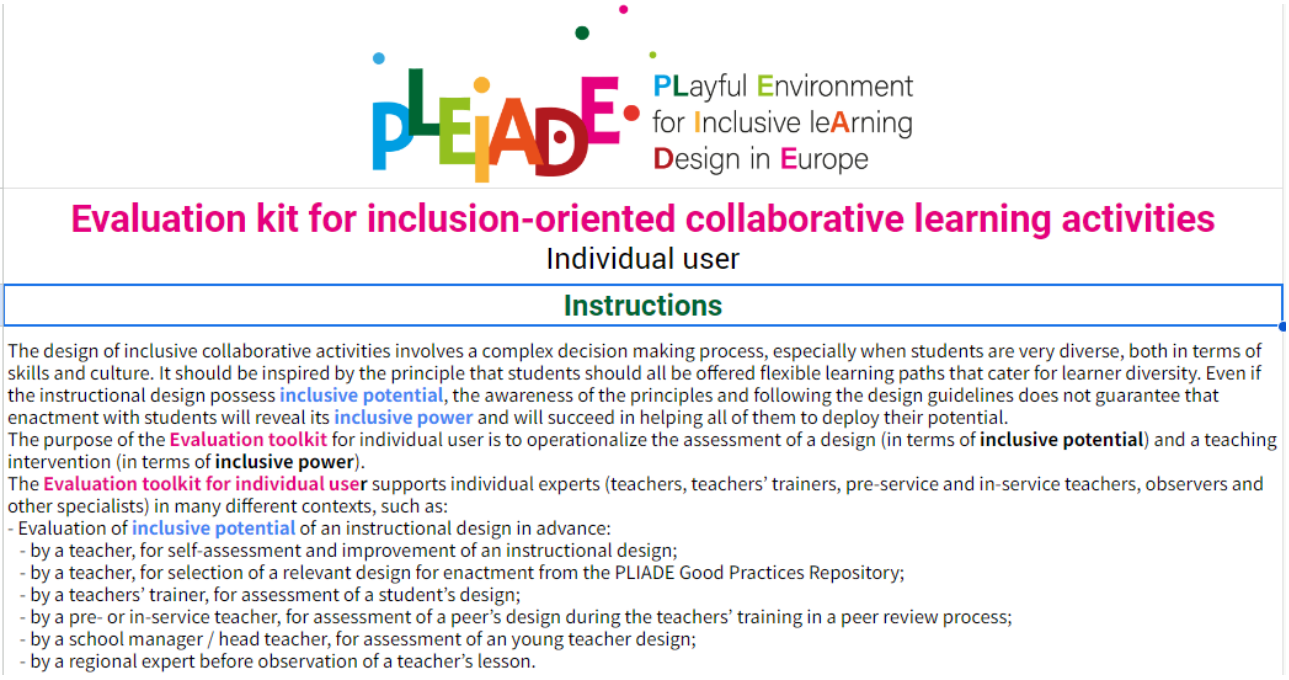


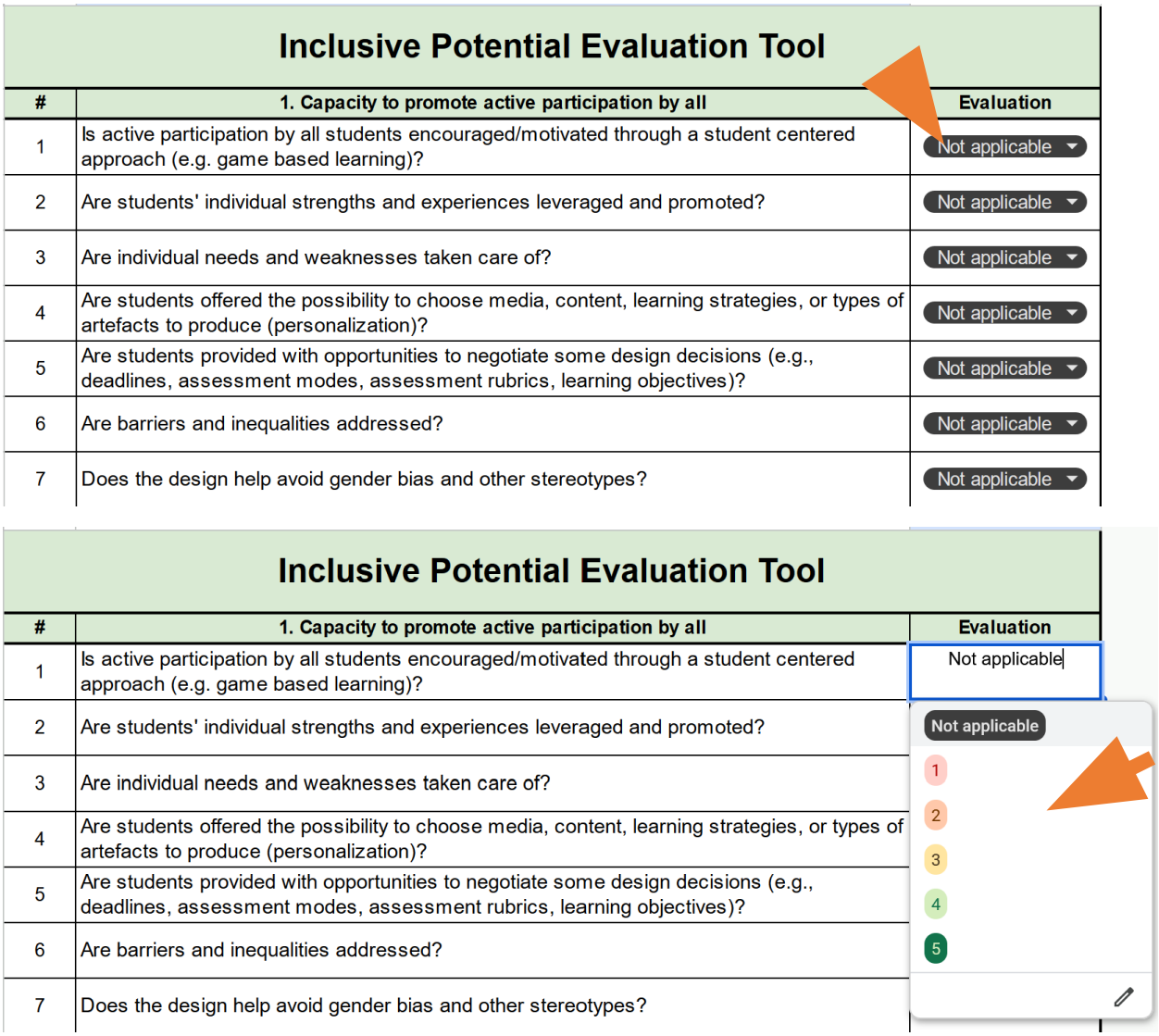
Figure 4. Instructions on how to use the kit for individual user

The Evaluation kit for individual user consists of seven sheets in total:

- Instructions
- Potential Evaluation
- Potential Result Chart
- Potential Suggestions
- Power Evaluation
- Power Result Chart
- Power Suggestions

In case the user would like to assess the inclusive potential of a collaborative learning design, he/she should go to the **Potential Evaluation** sheet; in case he/she would like to assess an enactment – to the Power **Evaluation sheet**.

Both sheets provide list of operationalized criteria and sub-criteria, developed through the Delphi study (Figure 5). Using the drop-down menu on the right, the



Inclusive Potential Evaluation Tool

#	1. Capacity to promote active participation by all	Evaluation
1	Is active participation by all students encouraged/motivated through a student centered approach (e.g. game based learning)?	Not applicable
2	Are students' individual strengths and experiences leveraged and promoted?	Not applicable
3	Are individual needs and weaknesses taken care of?	Not applicable
4	Are students offered the possibility to choose media, content, learning strategies, or types of artefacts to produce (personalization)?	Not applicable
5	Are students provided with opportunities to negotiate some design decisions (e.g., deadlines, assessment modes, assessment rubrics, learning objectives)?	Not applicable
6	Are barriers and inequalities addressed?	Not applicable
7	Does the design help avoid gender bias and other stereotypes?	Not applicable

Figure . Assessing the level of meeting each subcriterion

user should evaluate to what extent a given sub-criterion is met.

When the whole list is elaborated, the user should go to the **Potential Result Chart** or **Power Result Chart** sheet respectively. The chart provides a visual summary of the criteria. The red line shows the critical border, while the green one represents the result of the assessment (Figure 6).

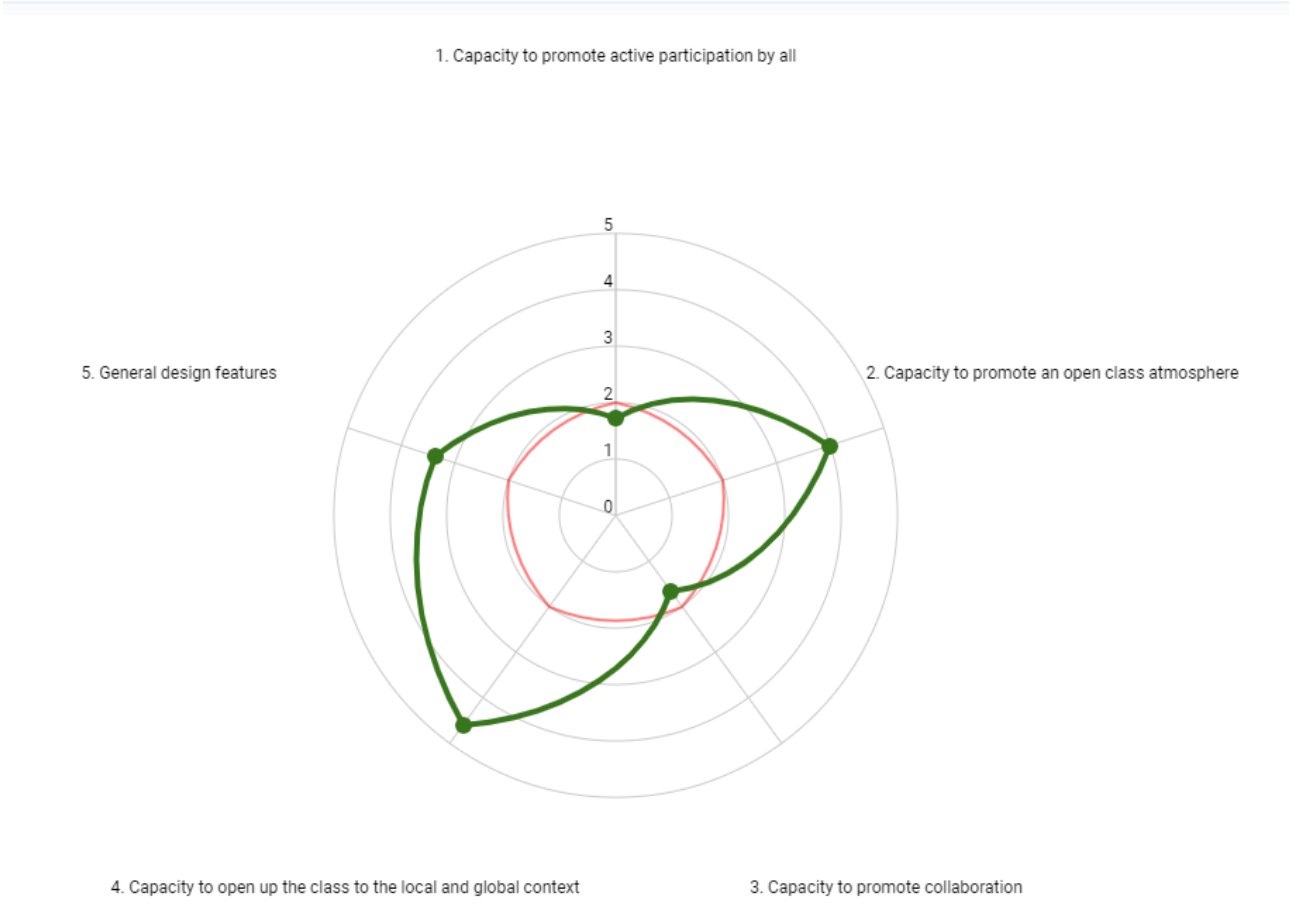


Figure 6. Example of a visual presentation of evaluation of a design that should be improved in relation to two criteria

The particular criteria, which should be improved, are also mention on the **Potential / Power Suggestions** sheet (Figure 7)

Figure 7. Suggestions sheet

7.2.4 Localization of the Evaluation Kit for individual user

The Evaluation kit for individual user is available in the following languages:

- English: https://docs.google.com/spreadsheets/d/1e0xJcl6ObUqU1RdrL-6kLmDo7eNRA3kncFzGXJZwF_g/copy

- Italian: https://docs.google.com/spreadsheets/d/1A607yHdGhpz6-VvitRPghQb9Cp6lMSghTGKOb8Jsi8I/copy
- Bulgarian: https://docs.google.com/spreadsheets/d/16WfdLFkPd4FZSGMq0T0_SZ9myH2DHEnWioOkI6EJWao/copy
- Greek https://docs.google.com/spreadsheets/d/1W_MAfJvzAtQ-yFSY44LZ7fRg0UzvT5hQe-zw34sTll4/copy

7.3 Evaluation tool for multiple users

7.3.1 Purpose

The purpose of the **Evaluation kit for multiple users** is to operationalize the assessment of a design (in terms of inclusive potential) and a teaching intervention (in terms of inclusive power). The **Evaluation kit for multiple users** could be used in pre-service teachers’ training, in continuous professional development (CPD) of teachers, or in ‘open lessons’ during regular teaching practice, for evaluation by a group of peers of:

- **inclusive potential** of a learning design;
- **Inclusive power** of teaching/learning process, achieved by the teacher’s intervention.

In this case, the evaluation process should be managed by a mentor, who sets up and distribute the form to the users, who will participate in the process.

The collected evaluation data are accessible only for the mentor. He/she has rights to share them or just the resulting chart / suggestions to the participants, author or publicly.

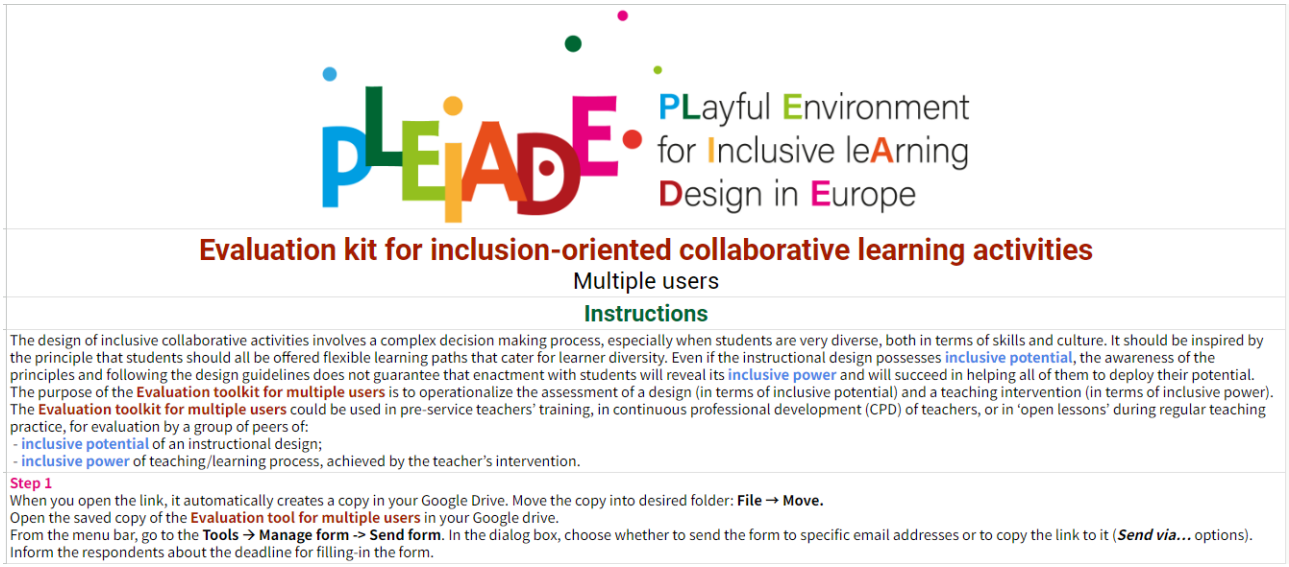
7.3.2 Access of the Evaluation kit for multiple users

The Evaluation kit for multiple users in English is accessible through the link: https://docs.google.com/spreadsheets/d/1Tq43e3810utId7_33ykxji4Y-cvOCMJ2f4TsyZQSdD4/copy or through the QR code:

The link asks the user (mentor) to create a copy of the original kit, so to have a private access to collected evaluation data.

7.3.3 Description of the Evaluation kit for multiple users

Similar to the Evaluation kit for individual users, after the copying, the kit loads the first sheet, containing instructions on how to use it (Figure 7).


PLayful Environment for Inclusive leArning Design in Europe

Evaluation kit for inclusion-oriented collaborative learning activities

Multiple users

Instructions

The design of inclusive collaborative activities involves a complex decision making process, especially when students are very diverse, both in terms of skills and culture. It should be inspired by the principle that students should all be offered flexible learning paths that cater for learner diversity. Even if the instructional design possesses **inclusive potential**, the awareness of the principles and following the design guidelines does not guarantee that enactment with students will reveal its **inclusive power** and will succeed in helping all of them to deploy their potential.
The purpose of the **Evaluation toolkit for multiple users** is to operationalize the assessment of a design (in terms of inclusive potential) and a teaching intervention (in terms of inclusive power).
The **Evaluation toolkit for multiple users** could be used in pre-service teachers' training, in continuous professional development (CPD) of teachers, or in 'open lessons' during regular teaching practice, for evaluation by a group of peers of:
- **inclusive potential** of an instructional design;
- **inclusive power** of teaching/learning process, achieved by the teacher's intervention.

Step 1
When you open the link, it automatically creates a copy in your Google Drive. Move the copy into desired folder: **File → Move.**
Open the saved copy of the **Evaluation tool for multiple users** in your Google drive.
From the menu bar, go to the **Tools → Manage form -> Send form**. In the dialog box, choose whether to send the form to specific email addresses or to copy the link to it (***Send via…*** options).
Inform the respondents about the deadline for filling-in the form.

Figure 8. Instructions on how to use the kit for multiple users

The copied Google sheet has assigned Google form. At this stage, if the mentor is familiar with Google forms, he/she is allowed slightly to edit the form – to add custom confirmation message, show/hide progress bar, etc. To edit the form he/she should go to the **Tools menu → Manage Form → Edit form**.

The form could be distributed among the respondents through the **Tools → Manage form → Send form** (Figure 9).

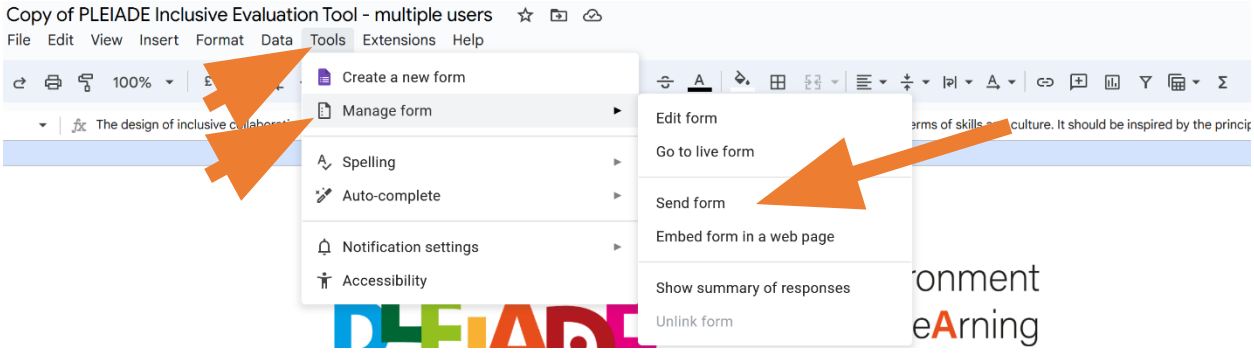


Figure 9. Form distribution

On the next step, the mentor can choose how to send the form – via email, providing a link through e-learning platform, or by embedding the form in a web site.

The respondents receive a ready for filling form (Figure 10).

Figure 10. Form, ready for filling

They can choose if to evaluate the inclusive potential of a design or the inclusive power of an enactment, and to continue to the relevant section (Figure 11).

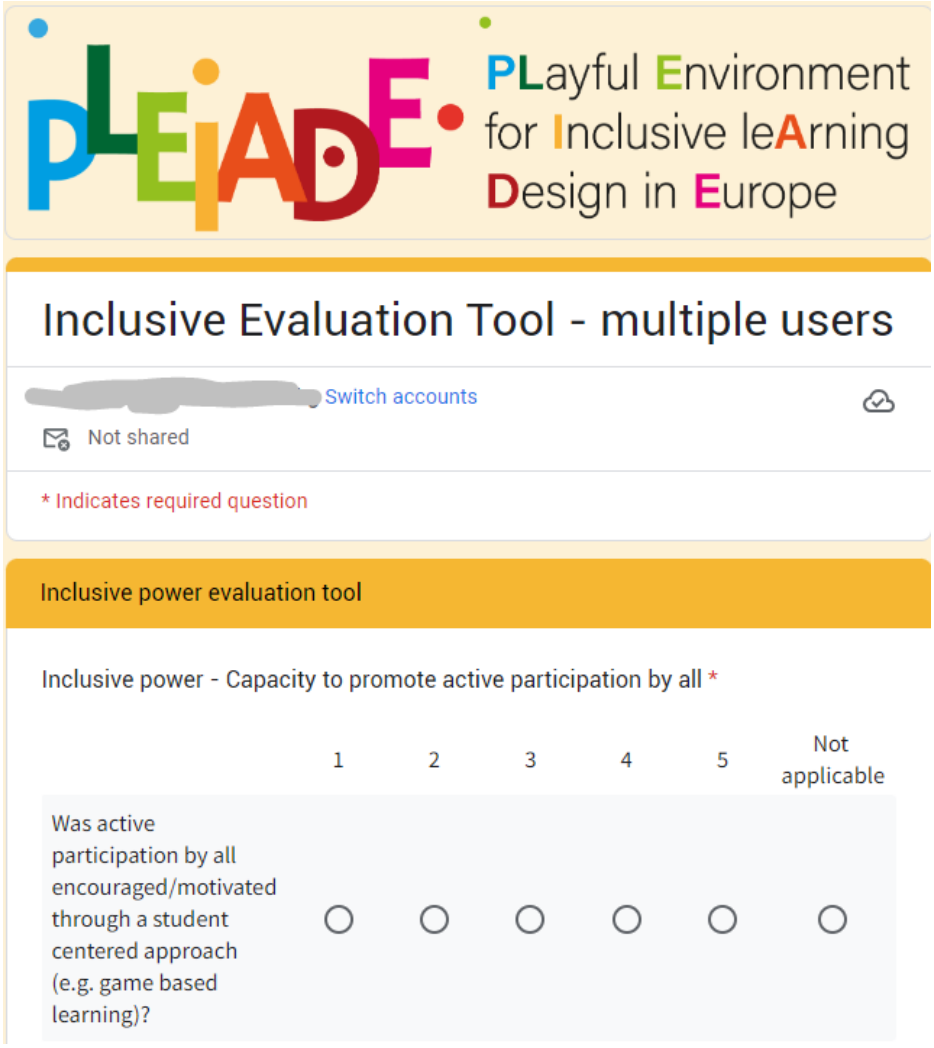


Figure 11. The multiple users form from the respondents' point of view

The answers are collected in a hidden sheet “Responses” in the main table (Figure 12).

	A	B	C	D	E
1	Timestamp	What tool do you want to	Inclusive potential - Ca	Inclusive potential - Ca	Inclusive potential - Ca
2	16/06/2023 17:09:00	Inclusive potential evalua	4	Not applicable	2
3	16/06/2023 17:20:36	Inclusive potential evalua	Not applicable	Not applicable	Not applicable

Figure 12. Data collected

The other hidden sheets – “Evaluation Potential” and “Evaluation Power” contain formulas for elaborating data and preparation the resulting chart and suggestions (Figure 13).

	A	B	C	D
1	**Question**		Inclusive potenti	Inclusive potenti
2	**Average Score**		4.00	Not applicable
3	**Weight**		5.00	5.00
4	**Calculated Weight**		5.00	0.00
5				
6	**Average by sections**	Average	Minimum	Suggestion
7	**1. Capacity to promote active participation by all**	3.15	2	
8	**2. Capacity to promote an open class atmosphere**	3.67	2	
9	**3. Capacity to promote collaboration**	3.45	2	
10	**4. Capacity to open up the class to the local and global context**	2.50	2	
11	**5. General design features**	1.00	2	Please revise yo

Figure 13. Data elaboration in hidden sheets

When the collecting data is complete, the mentor has the results.

The visualized graphs are accessible again on the sheets **Potential Result Chart** and **Power Result Chart (**Figure 14**)**. The sheets **Inclusive Potential Suggestions** and **Inclusive Power Suggestions** (Figure 15) provide again textual suggestions for improvement of the design or enactment. The result presentation is the same as in the Evaluation kit for individual user.

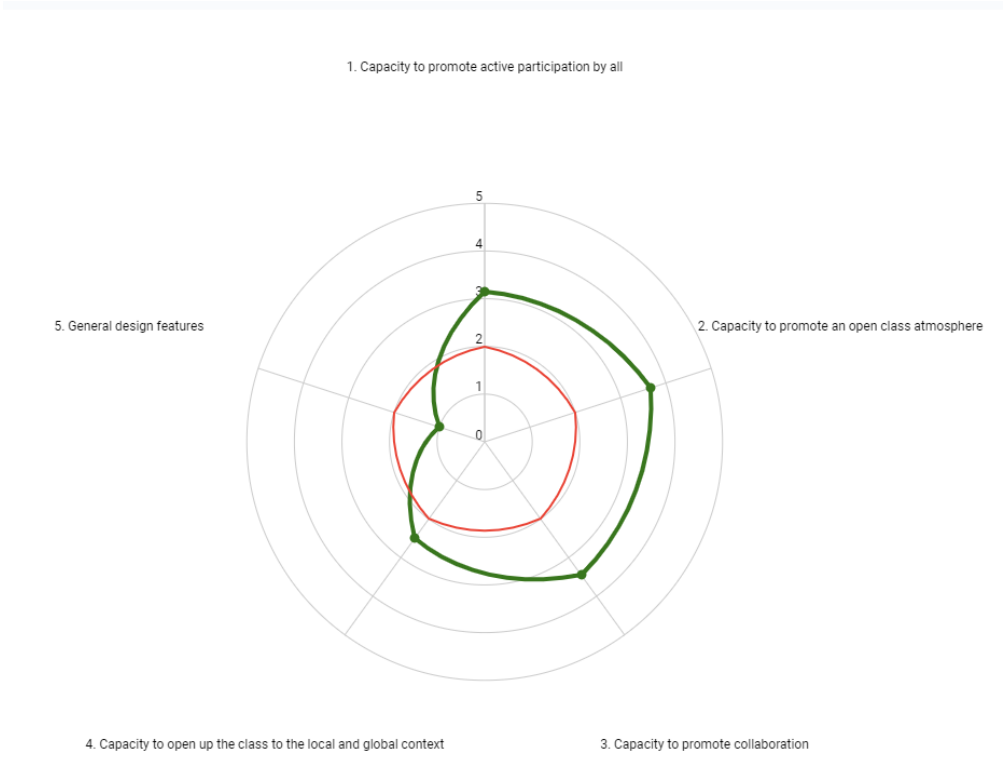


Figure 14. Resulting chart in the Evaluation kit for multiple user

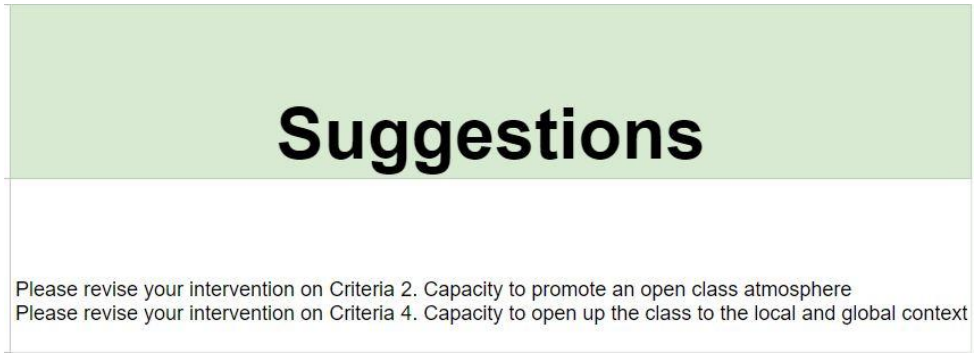


Figure 15. Textual feedback in the Evaluation kit for multiple users

If the mentor like to moderate a discussion or just guide the author on how to improve the design, he /she can go again to the live form: **Tools → Manage form→. Go to live form**, and discuss particular sub-criteria.

7.3.4 Localization of the Evaluation Kit for multiple users

The Evaluation kit for individual user is available in the following languages:

- English: https://docs.google.com/spreadsheets/d/1Tq43e3810utId7_33ykxji4Y-cvOCMJ2f4TsyZQSdD4/copy
- Italian: https://docs.google.com/spreadsheets/d/11xQqTV32OXlA1t3jJgHoxYpc0tUj1QxsJmRcvoCY9mI/copy
- Bulgarian: https://docs.google.com/spreadsheets/d/1CQ-WLKDXnUbIKGCE_pnEM208pedC83iM9oZCIqmMVuU/copy
- Greek: https://docs.google.com/spreadsheets/d/1oYSy6MpfXFZkQIorsdQPpONmGO7ItFbBOeGJWWqKyzk/ecopy

7.4 User Guide for using Evaluation kit

The complete User Guide for using the Evaluation kit is available at https://drive.google.com/file/d/1IiLLuhacOUvgIofN5btvRH3Nf0qbMENz/view?usp=sharing.

8 Validation of the evaluation toolkit

8.1 Aims of the validation process

The PLEIADE project aims to create a **Collection of good practices** (IO5) of tested scenarios for inclusive learning designs. During the previous stages of the project, CNR-ITD experts supported the work of teachers from Italy, Cyprus, Greece and Bulgaria to create inclusive learning designs. As a product of peer review and guidance from mentors, the designs are self-modelled. Some of the models have been tested in a real environment.

The role of the teams of researchers from Italy and Bulgaria is to support the teachers in their work and to critically comment on the proposed designs.

The main goal of the developed designs is **inclusive learning**. The framework for development and reflection emphasises **inclusiveness**.

The Sofia University team analysed the scientific literature in the field, the mentors’ comments and the results of the peer review and thus proposed a set of evaluation criteria.

The CNR-ITD team, based on the Delphi Study method, developed a questionnaire in collaboration with experts (panellists) and in parallel collected with teachers' contributions.

The aim of the next stage of the project is to validate the questionnaires and their usability as evaluation tools for inclusive potential and inclusive power of the developed learning designs.

The risk of using only qualitative analysis is in the formulation of the parameters. The questionnaires should contain clear and unambiguous questions for the interviewee so that the collected data can be representative, adequate, and appropriate. The joint work of experts from CNR-ITD and UniSofia made it possible to repeatedly analyse the criteria and sub-criteria, which reduces the possibility of bias.

The focus of the current section is the validation of the evaluation tools in respect of their **usability** and **reliability**. The process is carried out by monitoring and analysing of the **reactions** and **perceptions** of the teachers, while they evaluated the PLEIADE designs and implemented enactments. For this purpose, an additional questionnaire and a Think Aloud Protocol have been prepared.

According to the scientific literature, it is almost impossible to assess 100% the validity of a given instrument. Validity measures the accuracy, that means to what extent inclusive potential and power are considered.

8.2 Validation methodology

8.2.1 Types of validity concerned

8.2.1.1 Content validity and face validity

The main goal of the content validity is to check how well a given instrument gives an accurate idea of the considered characteristics (Rusticus, 2014). In this particular case – whether all the aspects about the inclusive potential and power are measured. A part of the observation is focused on how clear, unambiguous and applicable the questions are in terms of the response of the interviewees. The usual practice is face validity to be done by the researchers. In this case, the questionnaires are created by experts and UniSofia aims to validate the Evaluation tool, based on user behaviour. Therefore, observers incorporate **face validity** from the teachers' perspective.

8.2.1.2 Internal and external validity

The scenario creator evaluates his own design by the tool (self-checklist). This aspect of validation focuses on answering the questions to what extent the design creator judges that the preliminary criteria are met. At each stage, the formulation of the criteria and sub-criteria slightly differs from that of the previous ones due to the parallel Delphi study rounds. Here it is observed a look-back reflection, an analysis of what has been achieved in view of the expectations set in advance. It is interesting to follow the completion of the questionnaire by different participants who are co-authors of a given design, and the others. The completion of the questionnaire in the groups that have the opportunity to evaluate both potential and power is also indicative, although there is a risk of bias.

8.2.2 Reliability

The reliability was tested by two coefficients:

- **Inter-Rater Reliability (IRR)**
- **Cronbach’s Alpha**

Inter-Rater reliability is used to test the agreement between raters according to the provided scores. The simplest Joint-probability of agreement measure is used, which results in percentage of the raters agree on their decision (Uebersax, 1987):

$$IRR = \frac{Total\ Agreements}{All\ Possible\ Agreements} = \frac{Total\ Agreements}{Number\ of\ Raters \times Number\ of\ Items}$$

In the calculations of the *IRR* the meaning of *agreement* is ‘an agreement with the mean of scores of an elaborated item’. Each raters’ score rs_i is treated as ‘agreed’, if $\left|rs_i - mean\right| \leq 0.5$. The border of deviation of 0.5 is chosen as ordinal scales are used, and the difference between nodes in one unit.

According to Marques and McCall’s review (Marques & McCall, 2005), a rate in the range [50% - 90%] shows acceptable level of reliability.

In the PLEIADE case, IRR should ensure good raters’ agreement for untrained users of the kit. That is why it was measured during the first validation stage, while no one of the PLEIDE teachers was familiar with the tool.

The other calculated coefficient is **Cronbach’s Alpha** coefficient, which is used to verify the consistency of the items within the list (Barbera, Naibert, Komperda, & Pentecost, 2021). Cronbach’s Alpha is the average of all possible split-half reliabilities projected onto the number of measures in the scale (Zeller, 2005). A value, higher than or equal to 0.8 is interpreted as acceptable rate of reliability.

8.2.3 Usability of the tools

Usability of a tool affects users, administrators and researchers.

Users’ feedback is taken into consideration during the validation of the Evaluation tool. It is important to track the adequacy and correctness of the collected data and to what extent their format contributes to the researchers’ and practitioners' objectives, in this case – measurement of inclusive potential, inclusive power of a collaborative learning design.

For the purpose of the evaluation of the inclusive potential and analysis of the results with respect to inclusive power the following aspects were subject to usability validation:

- How clear are the instructions for using the tool;
- How easy is to fill it out on different devices, including a paper version;
- Portability;
- Administration time;
- Visibility and comprehensibility of the summary of the results – the chart;
- Problems noted using the tool.

Other aspects of the usability relate to the user-tool interactions:

- User behaviour based on written instructions
- Time for orientation in the tool structure and usage
- Alarm about any show-stoppers

- User understanding about the visual presentation of the assessment results with a view to inclusive potential and inclusive power
- Level of evidence about the link between summary of the results and answers of the questions, integrated in the Evaluation tool:
    - What changes could lead to improvement of the results?
    - Are the key aspects visible?
    - What is added value of the result-chart to the score of the check-list?

8.2.4 Process

Following the recommendations of Pedersen at all. (Pedersen, Emblemsvåg, Bailey, Allen, & Mistree, 2000), the Validation Square procedure is followed (Figure 16).

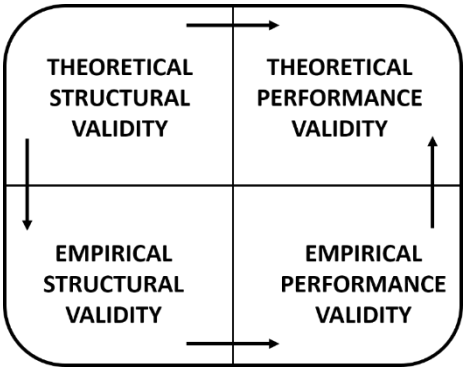


Figure . Validation Square CITATION Peder \l 1033 (Pedersen, Emblemsvåg, Bailey, Allen, & Mistree, 2000)

The validation stages were implemented in parallel with Delphi study. The results are used to validate the current Evaluation kit condition and to contribute to its further development. Qualitative and quantitative data is collected and elaborate as follow (Figure 17):

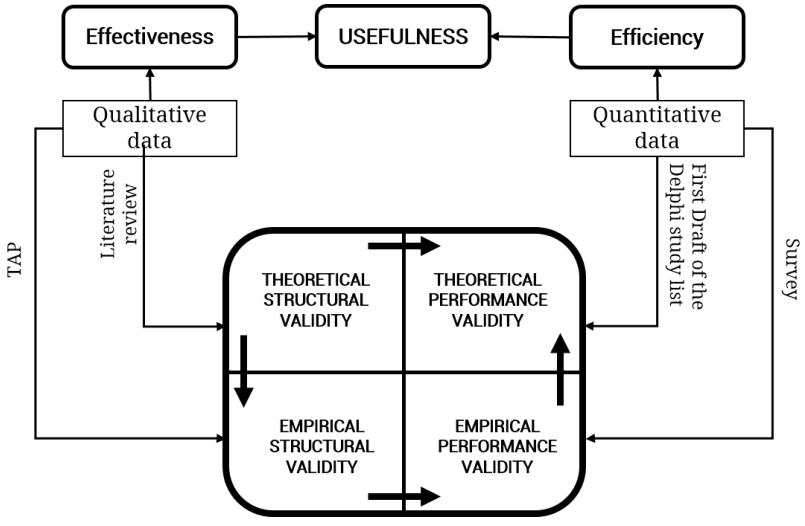


Figure 17. Validation Methodology - adaptation of the Design Method Validation (Pedersen, Emblemsvåg, Bailey, Allen, & Mistree, 2000)

1) **Theoretical Structural Validity**
    - Aim: Accepting the construct’s validity – theoretical literature review in advance, providing the main criteria for Delphi study; mapping the literature outputs to the end-users understanding and experience.
    - Qualitative data collected
    - Method and tools: literature review
    - Dates:
        i. April 2021 – December 2021 – literature review;
        ii. January 2022 – March 2022 – empirical study on PLEIADE designs and reviews.
2) **Theoretical Performance Validity**
    - Aim: Acceptance of constructs by the end-users, sub-criteria
    - Quantitative data collected
    - Methods and tools: first draft of questionnaire for Delphi study used for evaluation of selected designs and enactments
    - Dates: October 2022, SJSTE 3, Sofia,
3) **Empirical Structural Validity**
    - Aim: Accepting the structure of the Evaluation kit and the way of representation of the results
    - Qualitative data collected
    - Methods and tools: Think Aloud Protocol (TAP) while using the first prototype of the Evaluation kit for assessment of selected designs and enactments
    - Dates: March 2023, TPM 5, Athens
4) **Empirical Performance Validity**
    - Aim: Validation of the efficiency of the tool
    - Quantitative data collected
    - Methods and tools: The survey after using the Evaluation kit
    - Dates:
        - March 2023, TPM5, Athens – the first prototype of the Evaluation kit for individual user.
        - April 2023, Trani, as a satellite activity of the E3 – the Evaluation kit for multiple users.

Since the Evaluation kit development follows a design-based method, the validation process accompanied it, running in parallel with it, at three stages (Figure 18).

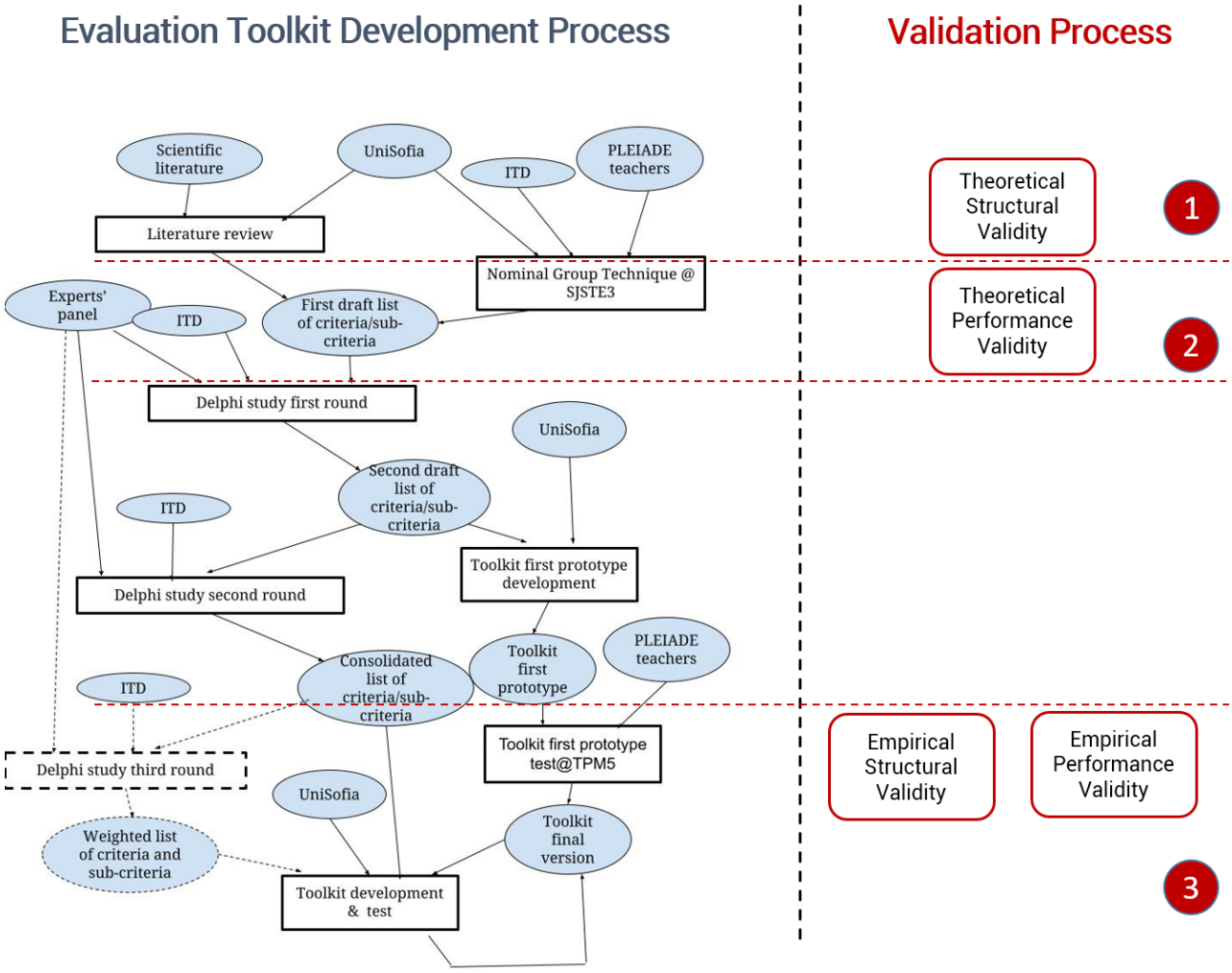


Figure 18. Validation process

The particular methods and tools used are described below, under each stage description.

8.3 Validation results

8.3.1 First test: reliability of the list of criteria and sub-criteria

The **Theoretical Performance Validity** is measured though the first test of the Evaluation kit, focused on the acceptance of the main criteria and subcriteria by the end users. It was implemented during the Third PLEIADE Short-term Joint Staff Training Event (SJSTE), 5 – 7 October, 2022, by 21 PLEIADE teachers. The validation was performed on the **First draft list of criteria and subcriteria** (**Error! Reference source not found.**), and it contributed also to the preparation of the list for the **First Round** of the Delphi study.

For the test needs, two of the designs – “Independence” and “Traditions”, were presented by their authors. The designs were chosen among those which were also

piloted in the period April – June 2022. The teachers who piloted the designs commented on their designs and enactment experiences, as well as the problems they met, and the possible changes that they plan for subsequent implementations, to improve the enactment.

After plenary presentations, the participants were divided into two groups, each one evaluating one of the designs and its enactment by the **First draft list of criteria**. In parallel, the teams of CNR-ITD and UniSofia took notes on their discussions and, at the end, presented a summary of subcriteria, which needed clarification.

A total of 21 PLEIADE teachers participated in this stage of the validation process. 7 of them used the First draft criteria list to evaluate the "Independence" design and enactment, and 13 – to evaluate the "Traditions" design and enactment.

For each criterion, the reliability is measured by two coefficients:

- **Inter-Rater Reliability (IRR)** – calculated as simple percentage agreement (Tong, Tang, Irby, Lara-Alecio, & Guerrero, 2020), where:
  - 1 (agreement) indicates that the difference between participant's score and the mean of a given sub-criteria is no more that 0.5 (as all the data is represented by Likert scales with nodes from 1 to 5, and StDev for the whole list is less than 0.5).
  - 0 (no agreement) indicates that the same difference is more than 0.5
- **Cronbach's Alpha** coefficient – to verify the consistency of the items within the list (Barbera, Naibert, Komperda, & Pentecost, 2021).

The Table 15 – Table 19 present the result of the evaluation of the inclusive potential & inclusive power of the presented designs through the First draft list of the criteria.

Table 15. Criterion 1 reliability

		Min	Max	Mean	StD
	CRITERION 1 **Coherence, completeness and clarity of the design**	**3**	**4**	**3.92**	**0.24**
1	aims are relevant for target needs	3	4	3.72	0.46
2	aims, learning objectives and contents are consistent with one another	3	4	3.67	0.49
3	assessment is aligned with learning objectives	2	4	3.56	0.62
4	the aims are in line with the curriculum	3	4	3.78	0.43
5	the design is clear	3	4	3.72	0.46

		Min	Max	Mean	StD
6	the teaching and learning approach is adequate to achieving the learning objectives	3	4	3.72	0.46
7	aims specifically target inclusion	3	4	3.67	0.49
8	the teaching and learning approach is adequate to achieving inclusion	3	4	3.78	0.43
9	contents promote inclusion (i.e. promote reflection about diversity, while encouraging reciprocal understanding)	3	4	3.61	0.50
10	the design is complete	3	4	3.94	0.24
	Number of valid judges)	**18**			
	Number of items	**10**			
	Cronbach Alpha	**0.65**			
	IRR	**0.72**			

Table 16. Criterion 2 reliability

		Min	Max	Mean	StD
	CRITERION 2 Capacity to promote active participation by all	3	4	3.93	0.27
1	active participation by all is encouraged/motivated	2	4	3.57	0.65
2	active participation by all is required	2	4	3.64	0.63
3	students' individual strenghts are leveraged	2	4	3.71	0.61
4	students individual needs and weaknesses are taken care of	3	4	3.71	0.47
5	students are offered choices (personalization)	3	4	3.79	0.43
6	barriers and inequalities are addressed	3	4	3.57	0.51
7	gender bias and other stereotypes are avoided	3	4	3.71	0.47
8	technological choices do not hinder access	3	4	3.93	0.27
9	assessment for learning is preferred to assessment of learning	0	4	3.00	1.36
	Number of valid judges)	**14**			

		Min	Max	Mean	StD
	Number of items	**9**			
	Cronbach Alpha	**0.70**			
	IRR	**0.75**			

Table 17. Criterion 3 reliability

		Min	Max	Mean	StD
	CRITERION 3 Capacity to promote a positive class atmosphere	4	4	4.00	0.00
1	all students are valued and respected	3	4	3.91	0.30
2	students support each other	3	4	3.82	0.40
3	achievements of all students are encouraged	3	4	3.80	0.42
4	the class atmosphere is relaxed	3	4	3.80	0.42
5	bullying or fights are prevented/avoided	3	4	3.56	0.53
6	the class atmosphere is playful	3	4	3.73	0.47
7	students and teachers emotions are positive	3	4	3.91	0.30
8	minorities are not isolated; there are no tight cliques	3	4	3.70	0.48
9	the teaching and learning approaches are adequate to foster social bonds	3	4	3.64	0.50
	Number of valid judges)	**14**			
	Number of items	**9**			
	Cronbach Alpha	**0.92**			
	IRR	**0.78**			

Table 18. Criterion 4 reliability

		Min	Max	Mean	StD
	CRITERION 4 Capacity to promote collaboration	3	3	3.00	0.00

		Min	Max	Mean	StD
1	collaborative learning is the core approach of the design	3	4	3.88	0.33
2	collaboration is mostly aimed at producing shared artefacts	3	4	3.71	0.47
3	interdependence and shared responsibility are fostered and rewarded	2	4	3.50	0.73
4	collaboration is incentivated and rewarded	1	4	3.56	0.81
5	tasks are conceived to promote collaboration	2	4	3.76	0.56
6	team size and composition facilitate collaboration	3	4	3.69	0.48
7	technological choices facilitate collaboration	3	4	3.76	0.44
8	timing is adequate to allow collaboration	2	4	3.75	0.58
9	the collaborative activities' design favour new contacts and interactions	0	4	3.59	1.00
10	assessment strategies summarize the accomplishments of both individuals and group	0	4	3.41	1.00
	Number of valid judges)	**17**			
	Number of items	**10**			
	Cronbach Alpha	**0.95**			
	IRR	**0.72**			

Table 19. Criterion 5 reliability

		Min	Max	Mean	StD
	CRITERION 5 **Capacity to open up the class to the local and global context**	3	4	3.56	0.51
1	contacts with other cultures are encouraged	2	4	3.78	0.55
2	understanding of different points of view is encouraged	2	4	3.67	0.59
3	The activity involves people or places outside the classroom, a local level	2	4	3.76	0.56
4	The activity involves people or places outside the classroom, at international level	1	4	3.53	0.87
	Number of valid judges)	**17**			
	Number of items	**4**			

		Min	Max	Mean	StD
	Cronbach Alpha	**0.65**			
	IRR	**0.79**			

The Cronbach **α** is greater than 0.80 for Critera 3 & 4. The coefficient values for other criteria are in range [65, 70], which is fully satisfying for the first list just before Round 1. The situation with IRR coefficient is similar – it’s value is at least 0.75 for Criteria 2, 3 & 5, and near to the border for the other two.

A StDev near or greater than 1 focuses attention on the formulation of the particular sub-criteria. Such values are signs for misunderstanding or different interpretation of the sub-criteria. Using the participants descriptive feedback, these sub-criteria were taken into account and reformulated, or represented by a set of new criteria before Delphi study Round 1. A category “Not Applicable” was also added , as some participants pointed out that some of the questions were not applicable to the presented designs.

The Appendix 3 contains a link to the anonymized primary data used in the validation process.

8.3.2 Validation of the Evaluation tool for individual user

After development of the prototype of the **Evaluation kit for individual user**, the end users tested it during the TPM5 in Athens, on March 13th, 2023.

8.3.2.1 Measuring the Empirical Structural Validity

Think Aloud Protocol (TAP) (Jääskeläinen, 2010), (Krahmer & Ummelen, 2004) is used for collecting qualitative data, providing information about the **Empirical Structural Validity**. The moderators collected aloud thoughts of the participants as well as description of unexpected behavior observed. The TAP contains two sections – the first one related to the Evaluation kit, and the second one – to the prototype of the Open Online Collection of Good Practices (IO5). The criteria and subcriteria, listed in the prototype of the Evaluation kit for individual user was prepared by CNR-ITD after the First Round of the Delphi Study, as it was provided for the Second Round.

The main observed categories, concerning the Evaluation kit, are:

- Understanding of instructions
  - Are the instructions clear for the participants
  - What questions, related to the instructions raised

    - Whether participants take unexpected actions
- Orientation
    - Are the participants easily oriented how to use the tool
    - What are the main stoppers in orientation
- Additional guidance needed
    - Do the participants ask for additional guidance
    - What kind of additional guidance they need
- Comments on the questions
    - Are the questions clear for participants
    - Do they have any doubts about what is the meaning of a given question
    - Do they have troubles where in the design or enactment description to look for reasons for evaluation on given subcriteria
- User understanding about the visual feedback
    - How clear is the visualization of the diagram on different devices
    - How the users interpret the diagram
- User understanding about the textual feedback
    - Is the feedback readable
    - Do it focus attention on the specific criteria and subcriteria
- Reflection: ideas for improvement in order to meet higher level of inclusiveness
    - What ideas the participants share in relation to the improvement of the evaluated design / enactment

The Think Aloud Protocol used, is provided in Appendix 4.

The whole group of 25 teachers from the PLEIADE schools were distributed in four heterogeneous groups – each group consisted of teachers from different countries, working on four differente designs and their enactments – “The world of languages”, “The divine intervention in people's lives in Euripides' Helen”, “Pythagorean Theorem”, “Individual Independence”. Among the members of each group, there was an author of the design or a teacher who has implemented it. This condition allowed: 1) the author to explain some curious points or answer questions, filling the gaps in information about enactment; and 2) to evaluate his own design / enactment. The descriptions of the designs and enactments were made available for the participants via the prototype of the Open Online Collection of Good Practices. As the Delphi study was still in progress and the refinement of the evaluation criteria and its operationalization was the responsibility of the experts’ group, the comments on the questions were used only for validation needs, but this time they did not lead to direct change of the criteria list. The focus was on the usability of the Evaluation kit as a whole, and any need for improvement.

The mentors of the four groups collected four TAPs. One of them was partially filled-in only in the part, dedicated to the evaluation of the prototype of the online collection. The mentor reported that the evaluation process went smoothly and he did not register any problematic situation. The summary from the other three TAP mentors is presented in the Table 20. In brackets, the number of protocols reporting the same / similar condition is shown.

Table 20. Summary of the TAP records

1. Understanding of the instructions	• Instructions are clear – (4) • Easy to follow the instructions (4) • Participants ask about the difference between *potential* and *power*. • **Need of install a spreadsheet app on the mobile or to login to Google Documents**
2. Orientation	• Easy, quick, no problems (4) • *My group provided guidance to another one* (comment by a mentor)
3. Additional guidance needed	• **Difficulty with making a copy of the tool (1)** • **A little difficulty accessing the tool via mobile (2)** • **Suggestion to provide QR code for mobiles (1)** • Some participant are stressed by the number of questions and asked to be shorten (1).
4. Comments on the questions	• It is better to use statements than questions (1 person) • **Proofreading of the questions needed (1)** • **Add a 'Not Applicable' option as some of the subcriteria are not applicable in a given context (2)**
5. User understanding about the visual feedback	• **A trouble in interpretation of the chart, especially what is the meaning of crossing the red and green lines (1)** • Easy, simple, readable (3) • Drop-down menus are not visible on some tablets
6. User understanding about the textual feedback	• Easy • Clear interpretation
7. Reflection: ideas for improvement to meet higher level of inclusiveness	• Splitting particular item into two different items • Recommendations from the group for improvement of the design / enactment provided (3)

All the four groups uploaded the successfully filled in Evaluation tools to the dedicated padlet: https://padlet.com/nnikolova73/pleiade-evaluation-tool-validity-wbpganmyoeacez4a.

The following improvements are implemented in the official version of the tool, taking into account the provided TAP results:

- Instructions about preliminary requirements for usage the kit provided
- The link provided creates a copy of the tool automatically.
- The provided initial instructions are more descriptive.
- QR codes are provided.
- Proofreading made by a native English speaker; the text of the subcriteria is corrected.
- ‘Not Applicable’ option added.
- Instructions on the radar chart interpretation added.
- Suggestion, related to the criteria and subcriteria (splitting questions, reformulation) are provided to the expert group for taking decision during the Round 2.

In addition, during the plenary discussion at the end, some of the participants shared inspirations to use the Evaluation kit during open lessons of a teacher, where to ask participating teachers to evaluate particular presented design and to use the results as a starting point for workshop.

Based on this idea, later an Evaluation kit for multiple users was developed.

8.3.2.2 Measuring the Empirical Performance Validity

The empirical performance validity was tested during the same meeting in Athens, and the same activity. After the evaluation of the provided designs and enactments, the participants were provided with a simple survey. They were asked to fill-in it individually.

The survey (Appendix 5) asked respondents to evaluate seven characteristics of the **Evaluation kit for individual user**.

- Time cost (is it time effective way for evaluation)
- Orientation how to use the tool (self-reflection)
- Easy to fill the scores
- Easy to find the results
- Easy to interpret the results and feedback
- Provides insights about other teaching approaches, ensuring higher level of inclusiveness

- Reusability – easy to copy the tool and to use it again on another design / enactment

The 3-levels evaluation scale was used, represented by emojis:

- 😊 – satisfied in consideration with the question;
- 😐 – neutral;
- ☹ - disappointed;

20 respondents filled-in the survey. For analysis needs, the emojis are coded with numbers as follow:

- 😊 – 1
- 😐 – 0
- ☹ – -1

Table 21 presents the summary of the results:

Table 21. Summary of the usability survey answers on the Evaluation tool for individual user

Design name	Count	Average of Time cost	Average of Orientation how to use the tool	Average of Easy to fill the scores	Average of Easy to find the results	Average of Easy to interpret the results and feedback	Average of Reusability
Individual Independence	6	0.67	0.67	1.00	1.00	0.50	0.50
Pythagorean Theorem	4	1.00	0.75	1.00	0.50	0.00	0.25
The world of Languages	5	0.80	0.00	1.00	-0.20	0.80	0.20
The divine intervention in people’s lives in Euripides' Helen	5	0.60	1.00	1.00	1.00	0.60	1.00
Grand Total	**20**	**0.75**	**0.60**	**1.00**	**0.60**	**0.50**	**0.50**
Min	0	0	1	-1	0	-1	-1
Max	1	1	1	1	1	1	1
Mean	0.75	0.60	1.00	0.60	0.50	0.42	0.50
StDev	0.44	0.50	0.00	0.68	0.51	0.61	0.69

All of the means are positive numbers, with standard deviations no more than 1. These values show strongly positive attitude to the usability of the Evaluation tool for individual user.

Primary data is available at
https://docs.google.com/spreadsheets/d/1FIjtRgGCwRR6JU4hqhhp5vCP94zHprxI/edit?usp=sharing&ouid=105181343480781209971&rtpof=true&sd=true

8.3.3 Validation of the Eval tool for multiple users

A teachers’ workshop, organized in the frame of the Italian Multiplier Event (E3) was used for testing the Evaluation kit for multiple users as a second stage of **Empirical Performance validity**. The workshop was held on April 18th, 2023. **27** teachers form Scuola secondaria di 1° grado "Rocca-Bovio-Palumbo" participated.

The authors of 2 designs - *Dalla parte giusta* and *Learn by playing*, presented their designs and key stages of their enactments. Next, the whole group was divided into two subgroups (A and B), each one assessing one of the presented designs and enactments by the use of the **Evaluation Kit for Multiple Users**.

Some of the teachers (5 teachers, who participated in the validation process in Athens) were familiar with the Evaluation kit for individual users and had expectations about the tools and representations of the results. For others the tool was completely new.

The observations by the mentors from CNR-ITD and UniSofia show quick orientation in the tool and short time (15 – 20 minutes overall) for filling it.

To measure the usefulness of the tool, the same emoji-survey was used.

The data (Table 22) show better orientation than in usage of the Evaluation tool for individual user, but harder to find the resulting chart. As the teachers are familiar with Google forms, they easily see how to replicate the tool and reuse it.

Table 22. Summary of the usability survey answers on the Evaluation tool for multiple users

Design name	Time cost	Orientation how to use the tool	Easy to fill the scores	Easy to find the results	Easy to interpret the results and feedback	Provides insights	Reusability
Learn by playing - B Count	11	11	11	11	11	11	3

Design name	Time cost	Orientation how to use the tool	Easy to fill the scores	Easy to find the results	Easy to interpret the results and feedback	Provides insights	Reusability
Learn by playing - B Min	0	1	-1	0	0	0	0
Learn by playing - B Max	1	1	1	1	1	1	1
Learn by playing - B Average	**0.73**	**1.00**	**0.55**	**0.55**	**0.91**	**0.91**	**0.67**
Learn by playing - B StdDev	**0.47**	**0.00**	**0.69**	**0.52**	**0.30**	**0.30**	**0.58**
Dalla parte giusta - A Count	16	16	16	16	16	16	13
Dalla parte giusta - A Min	0	0	0	0	0	0	0
Dalla parte giusta - A Max	1	1	1	1	1	1	1
Dalla parte giusta - A Average	**0.81**	**0.94**	**0.63**	**0.56**	**0.50**	**0.69**	**0.85**
Dalla parte giusta - A StdDev	**0.40**	**0.25**	**0.50**	**0.51**	**0.52**	**0.48**	**0.38**
Grand Count	27	27	27	27	27	27	16
Grand Min	0	0	-1	0	0	0	0
Grand Max	1	1	1	1	1	1	1
Grand Average	**0.78**	**0.96**	**0.59**	**0.56**	**0.67**	**0.78**	**0.81**
Grand StdDev	**0.42**	**0.19**	**0.57**	**0.51**	**0.48**	**0.42**	**0.40**

Primary data is available at https://docs.google.com/spreadsheets/d/1bH_G4jr-2sJMcvEi7N8ZMcZjY7MJHgi7/edit?usp=sharing&ouid=105181343480781209971&rtpof=true&sd=true

9 Conclusions and Future prospects

9.1 Conclusions

In conclusion, it is hoped that the evaluation of the inclusive potential and power of a learning design will be able to offer significant food for thought to both the designers and the teachers interested in using that design. It should also be noted that the main added value of a kit based on these criteria is diagnostic. In fact, the criteria and subcriteria identified through this Delphi study identify important characteristics that a design should have in order to be inclusive, and a low score regarding one of these characteristics is a litmus test of what should be improved to increase the inclusiveness of the instructional project.

The Evaluation kit for assessment of the **inclusive potential** of a collaborative learning design and the **inclusive power** of its enactment is easy to use, effective and efficient. It can be used in many different contexts, in relation with pre-service teachers’ training, continuous professional development (CPD) of teachers, supporting young teachers, and others. The kit consists of two tools – one for individual users and one for multiple users evaluating the same design. The list of criteria and sub-criteria is operationalized in two tools – for evaluation of the inclusive potential of a design and for assessment of evaluation power of enactment. Both tools could be used as lists of observable indicators, which, in parallel, could provide insights for better designs or activity management. The list for assessment of the inclusive power also provides the possibility for an enactment to be monitored in real time and important decisions for on-going changes of the classroom and lessons management to be taken in order to guarantee the inclusiveness of the learning process.

Untrained users are equipped with a well-illustrated User Guide.

As the Evaluation kit is well scientifically grounded, it can be used for research goals also.

The supporting document – the description of the design and delivery of development process, validation methodology and tools, are also valuable products of the Intellectual Output 4.

9.2 Future prospects

Although the Evaluation kit meets all of the requirements of the PLEIADE project, the Delphi study still is going on. The Third Round aim is to range the subcriteria according their importance and impact on the main criteria. The expected result is

weighted value for each sub-criteria. The refined kit by providing weights will be more precise in calculating the results.

The Evaluation kit could be improved also by adding functionality for identification of particular points where the design / enactment could be improved; preparation of more detailed suggestions on improvement; providing relevant materials for reading in order the author / teacher could be able to improve the design or enactment.

Appendix 1. Delphi Study on criteria for assessing inclusiveness of Learning Designs – First round

https://drive.google.com/file/d/1W_9K75O1GZcBDcq7clYBIgdcEbIBnFvj/view?usp=sharing

Delphi Study on criteria for assessing inclusiveness of Learning Designs

Dear member of the Panel of Experts of the PLEIADE Delphi Study,

thank you for accepting to participate in our study.
Aim of this Delphi Study is to define a set of criteria for assessing the inclusiveness of a Learning Design (aka lesson plan) produced by primary and lower secondary school teachers. By inclusiveness, we mean the capacity to generate inclusive learning processes, with particular focus on the inclusion of children with cultural, social or economic disadvantage, while also alleviating other exclusion issues due to belonging to minorities.

The results of the study will inform the development of one of the PLEIADE outputs: an evaluation toolkit that will be used to support the assessment of the designs produced by the PLEIADE teachers.
To implement the Delphi Study, we have set up a Panel of Experts (including you) who kindly agreed to provide their contribution by answering the questions of the following survey.
The survey was prepared by the PLEIADE partners on the bases of a number of criteria and sub-criteria that, according to the literature on Learning Design and Inclusive Education, should be fulfilled by good inclusive designs. The main criteria are reported below:
Capacity to promote active participation by all.
Capacity to promote a positive class atmosphere.
Capacity to promote collaboration.
Capacity to open up the class to the world.
Coherence, completeness and clarity of the design.
Through the survey, you will be asked to evaluate the importance of these criteria and their sub-criteria based on your professional experience. You will also be asked to suggest any new criteria or sub-criteria you believe are relevant for assessing the inclusiveness of a Learning Design.
The estimated time required to fill in the questionnaire is about 20-30 minutes.
As you may understand, your opinion is crucial in the process, and the researchers will weigh it carefully in combination with those of the other experts to produe a new survey, consisting in the second round of the Delphi Study. The process will go on until the new rounds will not yield new information.
We will be very grateful if you could fill in the survey at your earliest convenience, and

possibly no later than September 30th, 2022.
In case you need to get in touch with me, my contacts are below. Thank you so much again for your collaboration.
Donatella Persico
PLEIADE project coordinator [email: persico@itd.cnr.it]

1. **The information you will provide in this survey will be anonymized and presented only in aggregated form. Upon your consensus, your name will be listed in the project reports alongside those of the other members of the Panel. Your personal data will be managed in accordance with the PLEIADE project privacy policy.**
   Please confirm that you have carefully read the explanations concerning this research and the whole experimental procedure; that you have been informed about the aims and objectives of the research in question; that you have had the opportunity to ask questions about any aspect of the experimental procedure; that you have read the privacy policies and received satisfactory assurances about the confidentiality of the personal data collected by the research; that you are aware that you can withdraw at any stage of the study.

▢ Yes, I confirm

▢ No, I want to leave the survey

Processing of your personal data

2. **Please provide your consensus on the processing of your personal data according to GDPR, as part of this research in the terms and manner indicated above.**

▢ Yes, I consent

▢ No, I do not consent, and I want to leave the survey

Informant profile

Before we start collecting your opinions about criteria for assessing inclusiveness of a Learning Design, we need a little information about yourself and your knowledge in this area.

3. **Gender**

▢ Male

▢ Female

▢ prefer not to say

4. **Your age (in years)**

Only numbers may be entered in this field.

_____

5. **Country of your institution**

_____

6. **Your position**

Select all that apply

▢ researcher/academic

▢ teacher (higher education)

▢ teacher (upper secondary education)

▢ teacher (lower secondary education)

▢ teacher (primary education)

▢ policy maker

7. **Your specific field of expertise in inclusive education, if any**

Select all that apply

▢ intercultural education

▢ gender fairness

▢ inclusion of roma students

▢ inclusion of gifted students

▢ special needs education

8. **Please specify your disciplinary background:**

_____

Criterion 1: Capacity to promote active participation by all

9. **Importance of the sub-criteria**

Please rate the importance of each of the sub-criteria of the “Capacity to promote active participation by all” criterion for assessing the inclusiveness of a Learning Design.

(from 1=not at all important to 5=extremely important).
Select one per row

	1 (not at all)	2	3	4	5 (extremely important)	unsure/unclear
active participation by all is encouraged/motivated	○	○	○	○	○	○
active participation by all is required	○	○	○	○	○	○
students' individual strenghts are leveraged	○	○	○	○	○	○
students individual needs and weaknesses are taken care of	○	○	○	○	○	○
students are offered choices (personalization)	○	○	○	○	○	○
barriers and inequalities are addressed	○	○	○	○	○	○
gender bias and other stereotypes are avoided	○	○	○	○	○	○
technological choices do not hinder access	○	○	○	○	○	○
assessment for learning is preferred to assessment of learning	○	○	○	○	○	○

10. Completeness of the sub-criteria

Are there any other sub-criteria of the “Capacity to promote active participation by all” criterion that you deem relevant to assess the inclusiveness of a Learning Design? Please indicate below any missing sub-criteria you deem should be added:

Criterion 2: Capacity to promote a positive class atmosphere

11. Importance of the sub-criteria

Please rate the importance of each of the sub-criteria of the “Capacity to promote a positive class atmosphere” criterion for assessing the inclusiveness of a Learning Design.

(from 1=not at all important to 5=extremely important).

Select one per row

	1 (not at all)	2	3	4	5 (extremely important)	unsure/unclear
all students are valued and respected	○	○	○	○	○	○
students support each other	○	○	○	○	○	○
achievements of all students are encouraged	○	○	○	○	○	○
the class atmosphere is relaxed	○	○	○	○	○	○
bullying or fights are prevented/avoided	○	○	○	○	○	○
the class atmosphere is playful	○	○	○	○	○	○

students and teachers emotions are positive	○	○	○	○	○	○
minorities are not isolated; there are no tight cliques	○	○	○	○	○	○
the teaching and learning approaches are adequate to foster social bonds	○	○	○	○	○	○

12. Completeness of the sub-criteria

Are there any other sub-criteria of the “Capacity to promote a positive class atmosphere” criterion that you deem relevant to assess the inclusiveness of a Learning Design?
Please indicate below any missing sub-criteria you deem should be added:

_____

_____

Criterion 3: Capacity to promote collaboration

13. Importance of the sub-criteria

Please rate the importance of each of the sub-criteria of the “Capacity to promote collaboration” criterion for
assessing the inclusiveness of a Learning Design.

(from 1=not at all important to 5=extremely important).

Select one per row

	1 (not at all)	2	3	4	5 (extremely important)	unsure/unclear

Collaborative learning is the core approach of the design	○	○	○	○	○	○
Collaboration is mostly aimed at producing shared artefacts	○	○	○	○	○	○
Interdependence and shared responsibility are fostered and rewarded	○	○	○	○	○	○
Collaboration is incentivated and rewarded	○	○	○	○	○	○
Tasks are conceived to promote collaboration	○	○	○	○	○	○
Team size and composition facilitate collaboration	○	○	○	○	○	○
Technological choices facilitate collaboration	○	○	○	○	○	○
Timing is adequate to allow collaboration	○	○	○	○	○	○
The collaborative activities' design favour new contacts and interactions	○	○	○	○	○	○
Assessment strategies summarise the accomplishments of both individuals and group	○	○	○	○	○	○

14. Completeness of the sub-criteria

Are there any other sub-criteria of the “Capacity to promote collaboration” criterion that you deem relevant to assess the inclusiveness of a Learning Design?
Please indicate below any missing sub-criteria you deem should be added:

_____

_____

Criterion 4: Capacity to open up the class to the local and global context

15. Importance of the sub-criteria

Please rate the importance of each of the sub-criteria of the “Capacity to open up the class to the local and global context” criterion for assessing the inclusiveness of a Learning Design.

(from 1=not at all important to 5=extremely important).
Select one per row

	1 (not at all)	2	3	4	5 (extremely important)	unsure/unclear
Contacts with other cultures are encouraged	○	○	○	○	○	○
Understanding of different points of view is encouraged	○	○	○	○	○	○
The activity involves people or places outside the classroom, at local level	○	○	○	○	○	○
The activity involves people or places outside	○	○	○	○	○	○

the classroom, at international level	

16. Completeness of the sub-criteria

Are there any other sub-criteria of the “Capacity to open up the class to the local and global context” criterion that you deem relevant to assess the inclusiveness of a Learning Design? Please indicate below any missing sub-criteria you deem should be added:

_____

_____

Criterion 5: coherence, completeness and clarity of the design

17. Importance of the sub-criteria

Although this criterion doesn’t specifically pertain to inclusiveness, please rate the importance of each of its sub-criteria for assessing the overall quality of a Learning Design.

(from 1=not at all important to 5=extremely important).
Select one per row

	1 (not at all)	2	3	4	5 (extremely important)	unsure/unclear
the aims are relevant for target needs	○	○	○	○	○	○
aims, learning objectives and contents are consistent with one another	○	○	○	○	○	○

assessment is aligned with learning objectives	○	○	○	○	○	○
the aims are in line with the curriculum	○	○	○	○	○	○
the design is clear	○	○	○	○	○	○
the design is complete	○	○	○	○	○	○
the teaching and learning approach is adequate to achieving the learning objectives	○	○	○	○	○	○
aims specifically target inclusion	○	○	○	○	○	○
the teaching and learning approach is adequate to achieving inclusion	○	○	○	○	○	○
contents promote inclusion (i.e. promote reflection about diversity, while encouraging reciprocal understanding)	○	○	○	○	○	○

18. Completeness of the sub-criteria

Are there any other sub-criteria of the “Design coherence, completeness and clarity” criterion that you deem relevant to assess the overall quality of a Learning Design?
Please indicate below any missing sub-criteria you deem should be added:

_____

_____

Weighting criteria

19. Importance of the individual criteria

Please indicate how important you consider each of the 5 criteria proposed in this survey for assessing the quality and inclusiveness of a Learning Design.

(from 1=not at all important to 5=extremely important).
Select one per row

	1 (not at all)	2	3	4	5 (extremely important)	unsure/unclear
Capacity to promote active participation by all	○	○	○	○	○	○
Capacity to promote a positive class atmosphere	○	○	○	○	○	○
Capacity to promote collaboration	○	○	○	○	○	○
Capacity to open up the class to the local and global context	○	○	○	○	○	○
Design coherence, completeness and clarity	○	○	○	○	○	○

20. Further suggestions on criteria

Are there any other aspects that you deem relevant to assess the quality and inclusiveness of a Learning Design?
Please indicate below any missing sub-criteria you deem should be added:

_____

Survey conclusions

21. Suggestions concerning this survey

Please use the section below to provide any additional comment or suggestion concerning this survey.
Please write your answer here (optional):

---

22. If you agree to be further contacted about this study (for participating in the subsequent rounds and receiving the study report), please provide your name, surname and email below:

Write your name, surname and email below (optional):

---

23. Consensus to be listed as member of the Panel

Please, state if you agree that your name appears alongside those of the other members of the Panel in the reports about the study (data you are providing will be anonymized by the researchers and published only in aggregated form):

▢ Yes

▢ No

Appendix 2. PLEIADE Delphi study - Second round

https://drive.google.com/file/d/1pBm29Gb_pGYQwMY028wGVMVi6Twua6Br/view?usp=sharing

PLEIADE Delphi Second round

PLayful Environment for Inclusive leArning Design in Europe

Dear Member of the Panel of experts of the PLEIADE Delphi Study,

thanks again for your commitment to this study. This is a brief introduction to the Second Round Questionnaire, intended to provide some contextual information that should help you better understand its items.

As you may remember, **the aim of this Study is to define a set of criteria for assessing the inclusiveness of a Learning Design** (see our **glossary (https://docs.google.com/document/d/1Zcjg9GnFkXFmnyXvtwmr-_fBVzVYHDckKP6-jp32qkw/edit)**for a definition of the term) produced by primary or lower secondary school teachers. By inclusiveness, in PLEIADE, we mean the capacity to generate inclusive learning processes, with particular focus on the inclusion of children with cultural, social or economic disadvantage, while also alleviating other exclusion issues due to belonging to minorities. The results of the study will inform the development of one of the PLEIADE outputs: an evaluation toolkit that will be used to support the assessment of the designs produced by the PLEIADE teachers.

In the previous round of the survey, Panel Members rated the importance of five criteria for the inclusiveness of a learning design, along with their associated subcriteria. They also indicated new (sub)criteria, rewordings and suggestions for removal of the proposed criteria and subcriteria. By analysing the answers, we came up with a number of possible revisions to the list of criteria based on the panellists' suggestions. **In this second round of the Delphi study, we will show you the proposed revisions and ask you to express your agreement or disagreement with the suggested changes** to the list of (sub)criteria. Some of the respondents also raised questions about the intended meaning of some of the terms used in the questionnaire: it is for this reason we created a short **glossary (https://docs.google.com/document/d/1Zcjg9GnFkXFmnyXvtwmr-_fBVzVYHDckKP6-jp32qkw/edit)**of terms. Please refer to it when in doubt about how to interpret some terms.

In this Round, upon suggestion of some of the panellists, a new criterium is considered, and another has been slightly reworded. Thus, the following are the new criteria that form the basis for this round:

- Capacity to promote active participation by all.
- Capacity to promote an open class atmosphere.
- Capacity to promote collaboration.
- Capacity to open up the class to the world.
- General Design features.
- Contextual features and teachers' competences.

Appendix 3. Reliability of the list of criteria and sub-criteria

The full list of primary data is available at https://docs.google.com/spreadsheets/d/1NZU-ZGunnw8EJqJeIi-GGVx6dIMhXdYU3TfsGk5WkDE/edit?usp=sharing

Appendix 4. Think Aloud Protocol for validation of the Evaluation kit and Collection of Good Practices

Think Aloud Protocol – validation

Evaluation Toolkit

8. Understanding of the instructions	
9. Orientation	
10. Additional guidance needed	
11. Comments on the questions	
12. User understanding about the visual feedback	
13. User understanding about the textual feedback	
14. Reflection: ideas for improvement to meet higher level of inclusiveness	

Good Practices Repository

1. Orientation	
2. Finding the relevant design	
3. Use of search tool for searching other designs	
4. Readiness of the design	
5. Identification of the enactment description	
6. Rating	
7. Commenting	

Available also at:

- https://drive.google.com/file/d/1u3InUbd-UiZpcN6axxx229QcVPccMW-y/view?usp=sharing (one page)
- https://drive.google.com/file/d/1RMfS1OTy-P5K6CeZ4o1KvSUJSCCd5wO0/view?usp=sharing (two pages)

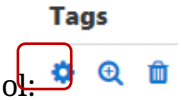
Appendix 5. Evaluation kit usability survey – Empirical Performance Validity

Design Name: Observer:

Evaluation kit – validation

Instructions:

1. Go to the PLEIADE Open online collection of good practices for inclusion (Demo): https://moodle.pleiade-project.eu/mod/data/view.php?id=348.
2. Split into two approximately equal sized subgroups.
3. Open the scenario, chosen by your facilitator. Use the lens tool:



4. Subgroup 1: Evaluate the design, using the Design Potential Evaluation Tool.
5. Subgroup 2: Evaluate the design enactment, using the Design Power Evaluation Tool.
6. Rate the design
7. Provide comments
8. Think / comment in the group aloud, while you are working.
9. Share your experience from evaluation process by checking the relevant emoji in the table below.

Design name: _____I am an author □
Subgroup: □ **Design** / □ **Enactment**

Characteristics			
Time cost (is it time effective way for evaluation)			
Orientation how to use the tool			
Easy to fill the scores			
Easy to find the results			
Easy to interpret the results and feedback			

Provides insights about other approaches, ensuring higher level of inclusiveness			
Reusability – easy to copy the tool and to use it again on another design / enactment			

Available also at:
https://drive.google.com/file/d/1bhX0NzzPhaQUTMMATWMZD9iK4UfRg2ym/view?usp=sharing

References

Barbera, J., Naibert, N., Komperda, R., & Pentecost, T. (2021). Clarity on Cronbach’s Alpha Use. *Journal of Chemical Education, 98*(2), 257–258. doi:https://doi.org/10.1021/acs.jchemed.0c00183

Booth, T., & Ainscow, M. (2011). *Index for Inclusion: Developing Learning and Participation in Schools.* Centre for Studies on Inclusiv.

Borko, H., Jacobs, J., & Koellner, K. (2010). Contemporary approaches to teacher professional development. *International encyclopedia of education, 7*(2), 548-556.

Bosse, I. K. (2015). Criteria for Designing Blended Learning Materials for Inclusive Education: Perspectives of Teachers and Producers. *Lecture Notes in Computer Science*, 3–14.

Chita-Tegmark, M. G. (2012). Using the universal design for learning framework to support culturally diverse learners. *Journal of Education, 192*(1), 17-22.

Davidson, G. R., & Carr, S. C. (2010). Forced migration, social exclusion and poverty: Introduction. *Journal of Pacific Rim Psychology, 4*(1), 1-6.

De Leeuw, R. R., De Boer, A. A., & Minnaert, A. E. (2019). Student voices on social exclusion in general primary schools. *Social Participation of Students with Special Educational Needs in Mainstream Education*.

Fabbri, L., Striano, M., & Melacarne, C. (2008). *L'insegnante riflessivo. Coltivazione e trasformazione delle pratiche professionali.* FrancoAngeli.

Floretta, V. (2021). Inclusive Education practices: an overview. *Academia Letters*, Article 1723. doi:10.20935/AL1723

Guskey, T. R. (2020). Flip the script on change: Experience shapes teachers' attitudes and beliefs. *The Learning Professional, 41*(2).

Herbert, S. (2011). *The inclusion toolkit.* Sage.

Jääskeläinen, R. (2010). Think-aloud protocol. *Handbook of translation studies, 1*, 371-374.

Jordan, A., & McGhie-Richmond, D. (2014). Identifying Effective Teaching Practices in Inclusive Classrooms. *Measuring Inclusive Education (International Perspectives on Inclusive Education), 3*, 133-162.

Krahmer, E., & Ummelen, N. (2004). Thinking about thinking aloud: a comparison of two verbal protocols for usability testing. *IEEE Transactions on Professional Communication, 47*(2), 105-117. doi:10.1109/TPC.2004.82820

Landeta, J. (2006). Current validity of the Delphi method in social sciences. *Technological Forecasting and Social Change, 73*(5), 467-482. doi:10.1016/j.techfore.2005.09.002

Lieberman, A., & Pointer Mace, D. H. (2009). The role of ‘accomplished teachers’ in professional learning communities: Uncovering practice and enabling leadership. *Teachers and Teaching: theory and practice, 15*(4), 459-470.

Loreman, T., Forlin, C., & Sharma. (2014). Measuring Indicators of Inclusive Education: A Systematic Review of the Literature. *Measuring Inclusive Education (International Perspectives on Inclusive Education), 3*, 165-187.

Marques, J., & McCall, C. (2005). The Application of Interrater Reliability as a Solidification. *The Qualitative Report, 10*(3), 439-462. From https://nsuworks.nova.edu/tqr/vol10/iss3/3/

Morrice, L. (2019). Abyssal lines and cartographies of exclusion in migration and education: towards a reimagining. *International Journal of Lifelong Education, 38*(1), 20-33.

Olimpo, G. (2011). Knowledge flows and graphic knowledge representations. *Technology and Knowledge Flow*, 91-131.

Passarelli, M., Dagnino, F. M., Ivanov, S., & Persico, D. (2022). Inclusive Pedagogy in Practice: The Point of View of Teachers. *ICERI2022 Proceedings* (pp. 5450-5458). IATED. doi:10.21125/iceri.2022.1332

Passarelli, M., Dagnino, F. M., Persico, D., Pozzi, F., & Nikolova, N. (2021). *Blended Teachers' Professional Development (TPD) pathway (PLEIADE Intellectual Output No. 1).* doi:10.17471/54009

Patton, K., & Parker, M. (2017). Teacher education communities of practice: More than a culture of collaboration. *Teaching and teacher education, 67*, 351-360.

Pedersen, K., Emblemsvåg, J., Bailey, R., Allen, J., & Mistree, F. (2000). Validating Design Methods & Research: The Validation Square. *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. 35142*, pp. 379-390. American Society of Mechanical Engineers.

Perla, L., & Martini, B. (2019). *Professione Insegnante.* Milano: Franco Angeli.

Persico, D., & Pozzi, F. (2011). Task, Team and Time to structure online collaboration in learning environments. *Journal on Educational Technology, 3*(1), 01-15.

Persico, D., Manganello, F., Passarelli, M., Pozzi, F., Nikolova, N., Stefanov, K., . . . Lonigro, M. (2023). Supporting teachers’ professional development on inclusive learning: a case study of an erasmus+ project. In L. Gómez Chova, C. González Martínez, & J. Lees (Ed.), *INTED2023 Proceedings - 17th International Technology, Education and Development Conference* (pp. 7456-7464). Valencia, Spain: IATED Academy. doi:https://doi.org/10.21125/inted.2023.2042

Persico, D., Milligan, C., & Littlejohn, A. (2015). The interplay between self-regulated professional learning and teachers’ work-practice. *Procedia-Social and Behavioral Sciences, 191*, 2481-2486.

Persico, D., Passarelli, M., Manganello, F., Gewerc Barujel, A., & Rodriguez Groba, A. (2023). The participatory dimension of teachers’ self-regulated professional learning about learning design: beliefs versus behaviours. *Professional Development in Education, 49*(2), 340-352. doi:10.1080/19415257.2020.1787193

Pleiade project. (2021). From Playful Environment for Inclusive learning Design in Europe (PLEIADE): https://pleiade-project.eu/

Pozzi, F., & Persico, D. (2013). Sustaining learning design and pedagogical planning in CSCL. *Research in Learning Technology*. doi:https://doi.org/10.3402/rlt.v21i0.17585

Pozzi, F., Ceregini, A., & Persico, D. (2016). Designing networked learning with 4Ts. *Proceedings of the 10th International Conference on Networked Learning*, (pp. 210-217).

Pozzi, F., Hofmann, L., Persico, D., Stegmann, K., & Fischer, F. (2013). Structuring CSCL through collaborative techniques and scripts. *Curriculum, Learning, and Teaching Advancements in Online Education*, pp. 113-122.

Pozzi, F., Manganello, F., & Persico, D. (2023). Collaborative Learning: a design challenge for teachers. *Education science*.

Rose, D. (2000). Universal design for learning. *Journal of Special Education Technology, 15*(4), 47-51.

Rusticus, S. (2014). Content Validity. (A. C. Michalos, Ed.) *Encyclopedia of Quality of Life and Well-Being Research*. doi:10.1007/978-94-007-0753-5_553

Tong, F., Tang, S., Irby, B. J., Lara-Alecio, R., & Guerrero, C. (2020). The determination of appropriate coefficient indices for inter-rater reliability: Using classroom observation instruments as fidelity measures in large-scale randomized

research. *International Journal of Educational Research, 99*, 101514. doi:https://doi.org/10.1016/j.ijer.2019.101514.

Trust, T., Krutka, D. G., & Carpenter, J. P. (2016). “Together we are better”: Professional learning networks for teachers. *Computers & education*(102), 15-34.

Uebersax, J. S. (1987). Diversity of decision-making models and the measurement of interrater agreement. *Psychological Bulletin, 101*(1), 140–146. doi:https://doi.org/10.1037/0033-2909.101.1.140

Zafirova-Malcheva, T., Antonova, A., Stamenkova, R., Nikolova, N., & Mihnev, P. (2022). Inclusive Potential of Learning Scenarios: Extracting Criteria for Evaluation. *INTED2022 Proceedings* (pp. 6101-6110). IATED. doi:10.21125/inted.2022.1550

Zeller, R. A. (2005). Measurement Error, Issues and Solutions. In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement* (pp. 665-676). Elsevier. doi:https://doi.org/10.1016/B0-12-369398-5/00109-2