BUS LEAGUE

Dedicated to stimulate demand for sustainable energy skills in the construction sector

www.busleague.eu

Report:	**D5.4 Evaluation Report of the Educational Perspective and the Development of the Evaluation Framework - Evaluation123**
Prepared by:	Sebastian Dennerlein & Maaike Endedijk, UT
Date:	31-01-2023
Partners involved:	ISSO, PF, EnE, IVE, LIT, AEA

This project has received funding from the European Union's h2020 framework programme for research and innovation under grant agreement no 892894.

TABLE OF CONTENTS

CHANGE RECORDS

Version	Date	Author	Changes
Version 1	27-10-2020	Miguel Angel	----------
Version 1	02-06-2021	Loes de Jong & Maaike Endedijk	First draft to literature study
Version 2	24-08-2021	Loes de Jong	Added scientific body
Version 3	05-09-2022	Sebastian Dennerlein	First Draft of complete structure and design trajectory
Version 4	16-01-2023	Maaike Endedijk	Feedback on deliverable
Version 5	06-06-2021	Sebastian Dennerlein	Completion of deliverable contents
Version 6	09-02-2023	Dirk de Wit	Feedback on deliverable
Version 7	14-02-2023	Sebastian Dennerlein, Sandra Schele & Maaike Endedijk	Finalisation and formatting of version for internal review
Version 7.1	16-2-2023	Dirk de Wit	Minor remarks
Version 7.2	27-2-2023	UTwente	Elaboration minor remarks

SUMMARY

The BUS League consortium has conducted upskilling interventions or trainings to address the stimulated demand for a skilled workforce in the energy transition. To design interventions for these activities and evaluate their effectiveness, it is important to identify what learning outcomes want to be addressed and what learning outcomes have been achieved.

Aside from their positive impact on self-regulated professional learning in terms of the informed selection of trainings and design of suitable trainings, only the focus on learning outcomes allows developing evaluations for acquired knowledge and skills, respecting the trade-off between evaluation benefits and efforts, and sharing resulting evaluation practices.

In the guidance and analysis of effectiveness and timeliness from the educational perspective, not only the objective learning outcomes and triggers are considered, but also the experienced outcomes: how were participants motivated to engage, did they experience the intervention help them to progress and how did they integrate what was learned in their current practice?

The subsequent report consists out of the following three chapters to explain the point of departure for effective evaluations, the co-design of support and the achieved insights:

1) Overview of how upskilling interventions currently record learning outcomes and what research recommends in terms of understanding and measuring upskilling.
2) Guide the design and implementation of evaluations of upskilling interventions to support training providers, including the presentation of a toolbox to do these evaluations, the EVALUATION123.
3) Demonstrate evaluation designs and insights of selected interventions to inspire future evaluations in the energy transition.

1 EDUCATIONAL PERSPECTIVE ON EVALUATION

BUS League is strengthened by anthropological and educational researchers in order to understand and assess impact of upskilling interventions or trainings, and to subsequently improve on the learning interventions and materials. These efforts aim at understanding training benefits for stimulating demand and optimising learning transfer that is essential to pave the way for a successful energy transition.

The educational perspective in BUS League provides guidance in evaluating the effectiveness of the upskilling interventions. The co-design of the guidance is informed by an exploration of the status quo in practice and literature. Its application is demonstrated in the BUS League project to inspire future upskilling and evaluation ambitions in the energy transition.

1.1 MEASURE EFFECTIVENESS AND TIMELINESS FROM EDUCATIONAL PERSPECTIVE

A multitude of upskilling interventions are created and implemented as part of the energy transition and the BUSLeague project. To evaluate the effectiveness and timeliness of these activities, it is important to identify what learning outcomes have been defined and evaluate them.

Evaluations are designed to assess the achievement of specific learning outcomes and it can be studied how they have been integrated in practice as well as what is missing for a meaningful transfer of learned knowledge and skills. To better understand how we can increase the demand for upskilling in the energy transition, it is necessary to also understand the motivational triggers and boundaries that participants experienced when engaging in the various interventions. We acknowledge that these elements are related: i.e., successful upskilling experiences that contribute to improvement of daily practice will stimulate the demand for additional upskilling of themselves and their colleagues.

Therefore, evaluations have to measure objective learning outcomes, but consider the experienced subjective learning outcomes and triggers as well: e.g., *how were participants motivated to engage, did they find the intervention as useful, did they feel they progressed on their learning goals and how did they integrate what was learned in their current practice?*

The **Task 5.3 of the Work Package 5 focused on the design of a step-by-step guidance for the evaluation of upskilling interventions in collaboration with the partners of BUS League**. The outcomes of T5.3 also showcase successful evaluation plans (also called evaluation design), together with respective insights about the domain. This results in the following sub tasks that are presented in the three main chapters of the Deliverable 5.4:

- **Chapter 1:** EXPLORING STATUS QUO IN PRACTICE AND LITERATURE
  Overview of evaluation design of the trainings in the BUS League and elicitation of state-of-the-art from the educational perspective in a literature study.
- **Chapter 2:** PROVIDING GUIDANCE FOR EVALUATING UPSKILLING (EVALUATION123)
  Proposal of guidance in assessing the impact of upskilling interventions for diverse evaluation goals and sketching the next steps in terms of learning analytics and paths.
- **Chapter 3:** DEMONSTRATING EVALUATION DESIGNS AND INSIGHTS
  Depiction of prominent evaluation context, designs and insights in energy transition and future plans of selected upskilling interventions.

1.2 RESEARCH APPROACH AND WORK PLAN

To collaboratively design (co-design) useful and feasible guidance for evaluating the effectiveness and timeliness of upskilling interventions in the energy transition, and in particular of the BUS League project, we implemented a Design Research Project from the Educational Perspective.

For this purpose, we drew upon the Generic Model for Conducting Design Research in Education (McKenney and Reeves, 2018). This model posits three core phases "analysis and exploration", "design and construction", and "evaluation and reflection" – in an iterative, responsive, and flexible structure. It also integrates the research and design processes with theoretical scientific knowledge and practice.

Table 1 gives an overview of the activities that were carried out in the three phases of educational design research. Details of the meetings conducted alongside the design research trajectory can be found in the Chapter 3 and the Appendix 7.1.

Design Phase	Research and Co-Design Activities	Planning
1. Analysis & Exploration	- Exploration of existing practices regarding upskilling activities and evaluation of effectiveness of BUSLeague partners by means of: 1. Survey research 2. Interview research - Literature study on effectiveness of upskilling activities	M1 - M12 (Sep 2020 - Aug 2021) **MS17: Plan for T5.3**

2. Design & Construction	- Development of initial version of evaluation framework (& rubric) based on insights from exploration phase
	- Co-Design evaluation framework and toolset for upskilling activities by means of
	• 4 pilots
	  ○ 2 with AEA (AT)
	  ○ 1 with Bauhaus (ES) &
	  ○ 1 with IVE (ES)
	• 2 consortium workshops (online)
	- Resulting "EVALUATION123" includes a manual for guiding interventions, a tool for formulating learning goals, directly applicable online questionnaire packs and a decision tree for planning evaluations
		M13 – M22 (Sep 2021 – June 2022)
		MS24: Evaluation for T5.3

3. Evaluation (& Implementation)	- Testing evaluation framework and toolset by means of a beta-testing at 5th consortium meeting: • Developing evaluation plans for various upskilling interventions in the consortium • Collecting of quantitative and qualitative feedback to understand effectivity of the designs and finalising then	M23 – M26 (July 2022 – October 2022) **MS24: Evaluation for T5.3**

Table 1: Work plan

2 EXPLORING STATUS QUO IN PRACTICE AND LITERATURE

2.1 UNDERSTANDING EXISTING UPSKILLING INTERVENTIONS - SURVEY RESEARCH

On December 21th 2020, BUSLeague interlocutors were invited to complete a survey on upskilling interventions for Task 5.3: Measure Effectiveness and Timeliness from Educational Perspective (M3-M22). All interlocutors except AVE have completed the survey. A total of 11 (clusters of) interventions have been reported by the interlocutors from BG (3), NL (5), IE (2), AT (0), and SP (1). Characteristics of the interventions are shown in Table 2.

Characteristics of interventions per country											
	BG			NL					IE		SP
	1. NZEB Training	2. Individual upskilling courses at fairs	3. Online training for homeowners' associations	1. EE-Skills	2. Girl's day events	3. Upskilling in DIY-stores on awareness	4. Cooperation with Breman academy	5. Promotion of the BUS-app	6. NZEB trainings	7. Retrofit skills	8. Bauhaus
Target group = workers	x	x		x		x	x	x	x	x	x
Currently running			x					x	x	x	
Starting within 1 year	x	x		x		(x)	x				X
Training/course for professionalization	x	x	x	(x)		x	(x)		x	(x)	x
Information available in Moodle	x			x					x		x
Information available in BUS-app				x			x			x	
User review	x			x	x		x		x	x	x

Table 2: Characteristics of Interventions Per Country

Note. Information is incomplete when a cross is in brackets (x).

2.2 EXPLORING EVALUATION NEEDS OF INTERVENTIONS - INTERVIEW RESEARCH

The desk research and the overview on implementation progress (T5.1) did not provide sufficient information to explore the existing practices regarding upskilling interventions and evaluation of effectiveness of BUSLeague partners. Therefore, all interlocutors were invited for short online interviews in May 2021, to provide additional information on the upskilling

interventions (planned to be) running and the type of questions they have in relation to researching educational effectiveness. Below, the findings from the interviews are summarised.

2.2.1 *Main upskilling interventions*

- BUS app for craft workers
- F2F (Face-to-Face) training for craft workers
- Micro learning modules, e-learning (through app or platform) for craft workers
- F2F training for staff DIY (Do It Yourself)/hardware stores (to be able to sell in the stores)
- Short electronic video clips for staff of DIY/hardware stores (to be able to sell in the stores)
- F2F training for energy advisors that go by family houses to seal

2.2.2 *Questions/needs i.r.t. researching educational effectivity*

- What is the effectiveness of upskilling interventions in terms of new knowledge, skills, improved learning attitude/motivation, applicability to daily work of participants?
- How can measuring/showing the educational effectiveness of upskilling interventions increase the demand side (e.g. upskilling is included as a requirement in procurement by grant provider)?
- ISSO would like to have in the end a work schedule/bingo sheet with overview of design principles for upskilling interventions and tips on how to evaluate the effectiveness (e.g. make an experimental and control group)

2.2.3 *Ideas/comments i.r.t. upskilling interventions*

- What are important characteristics of upskilling interventions (advice from interlocutors): user friendliness (e.g., attractive, direct, short), up to date information, simple language, access to WIFI, learning gains also visible to employers?
- In general, construction companies are not motivated to send they employees to training; craft workers are not motivated for learning (do not see the added value, they don't think they need it); blue collar workers prefer learn by doing; white collar workers like to reflect on work more (genuine interest); need for flexible/blended

2.3 LITERATURE STUDY ON THE EFFECTIVENESS OF UPSKILLING INTERVENTIONS

With upskilling we refer to all forms of learning for work. Upskilling takes various forms, ranging from formal training to informal discussions with colleagues (Littlejohn, 2017). In-service **continuous professional development or learning are other labels** used to denote the upskilling after one has graduated from formal vocational or higher education (e.g. Berndt et al., 2017; Bluestone et al., 2013; Cant & Levett-Jones, 2021). Broadly two approaches to upskilling can be distinguished, including the conventional approach and the increasingly popular workplace learning approach (Cheng & Hampson, 2008).

2.3.1 *Two approaches to understanding upskilling*

There are two approaches that can help to understand upskilling, the conventional approach and the workplace learning approach. Both approaches help us to define effectiveness in upskilling and infer respective measurements.

Conventional approach to upskilling

The *conventional approach* to upskilling is concerned with **formal training that takes place in isolation from workers' daily work, i.e. learning off-the-job**. Examples are courses, conferences, lectures, workshops, seminars, symposia (e.g. Berndt et al., 2017; Saks & Belcourt,

2006). Gao et al. (2019) distinguish between traditional tools (TT, e.g. as lecture, toolbox talk, handout, audio-visual material (e.g., video demonstration), computer-based instruction, and hands-on training) and computer-aided technologies (such as Serious Games (SG), Computer-generated Simulations (CGS), Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR)) for (health and safety) training in the construction sector. The authors identified several benefits of computer-aided material over traditional tools: representing the actual workplace situations, providing text-free interfaces, and having better user engagement.

Furthermore, the conventional approach to learning foresees the definition of specific learning outcomes, so called learning goals and objectives; on their basis, learning materials and activities can be designed systematically as well as learners assessed fairly (Anderson & Krathwohl, 2001; Oliver et al., 2004; Hattie, 2008). Despite being used interchangeably (Marzano, 2009), the terms learning goals and objectives have adopted distinct conceptualizations. In general, learning goals are used to refer to the broader, general possible intended outcomes of a learning activity (Fessl et al., 2021; Marzano, 2010; Mcnall, 2018), whereas learning objectives describe more specific, observable or tangible outcomes, expressed in behaviours resulting from a learning experience (McNall, 2018; Schiekirka et al., 2013).

However, the formulation of learning goals and objectives and their translation into clear, effective statements requires effort and attention from designers of instruction and instructors (De Long et al., 2005). In a study on the impact of explicating learning goals, Fessl (2021) found that about 50% of goals set by university teachers were poorly formulated, i.e., they contained redundant or unnecessary contents for the intended final learning outcomes. Concurrently, it has also been noted that different stakeholders may have distinct, and sometimes divergent understandings of the learning goals and objectives of a program (McNall, 2018).

Studies that prescribe a systematic for the formulation of learning goals and objectives generally agree that effective learning goals consist of a clear statement, declaring what is expected to be performed by an individual at a determined condition. Most studies also adopt a taxonomy (Ferguson, 1998; Fessl et al., 2021; Marzano, 2010; Wei et al., 2021) for the classification of level of learning and selection of one measurable item. Bloom's Taxonomy (Anderson & Krathwohl, 2001) is the most widespread tool used, either in its full format (remember, understand, apply, analyse, evaluate and create) or in adaptations derived from it (Fessl et al., 2021; Marzano, 2010).

Workplace learning approach

The *workplace learning approach* to upskilling is concerned with **processes that occur within specific organisational contexts** (Garavan, et al., 2002; Park et al., 2021), **i.e. learning on-the-job**. According to Tynjälä (2013), forms of workplace learning are needed that enable people to engage in transformative and innovative rather than in reproductive learning, and in networked and social learning rather than in individual learning, as well as in ethical and value conscious rather than 'value-free and objective' learning.

Whereas formal training often takes place alongside predefined curricula and resulting in certificates, workplace learning is characterised by self-directed exploration and social exchange that is tightly connected to the processes and places of work. Unlike learning in a formal setting, workplace learning is driven by personal interest or need or problem originating in the work context (Ruiz-Calleja et al., 2017). Workplace learning activities concern unintentional, intentional, informal, non-formal, and formal learning through doing the job itself, reflecting and evaluating one's own work experiences, collaborating with others, tackling new challenges and so on (Tynjälä, 2013).

Computer-aided technologies (e.g. Gao et al., 2019), MOOCS (e.g. Pilli & Admiraal, 2017; Wang, 2003; Wei et al., 2021), and e-learning (Chao et al., 2009; Tzeng et al., 2007) offer professionals the flexibility to decide when and where participate in (online) upskilling interventions, therewith blurring the boundary between the conventional and workplace approach to learning. Namely, e-learning offers professionals the opportunity to upskill without the constraint of time and space, especially in an asynchronous distance e-learning system. Besides learner flexibility, e-learning is a burgeoning educational and training tool for organisations because of the advantage

of cost saving and reusability. However, more attention is needed for professional networking opportunities or communities of practice because that is generally not part of e-learning (Berndt et al., 2017).

2.3.2 Defining effectiveness of upskilling

Gao et al. (2019) point to the **importance of understanding the effectiveness of upskilling interventions**. In the construction sector, especially computer-aided technologies have gained increased attention from many construction companies all over the word while researchers warn that the adequate evidence of new technologies' educational effectiveness is lacking (e.g. Gao et al., 2017; Choi et al., 2017). Different research approaches to the effectiveness of upskilling exist (see Appendix7.1.2 for an overview over proposals for measuring educational effectiveness).

Conventional approach to learning

From the conventional approach to learning, **effectiveness can be defined as the extent to which upskilling interventions yield desired outcomes** (Ho & Dzeng, 2010; Gao et al., 2019). Hattie (2008) states that having clear goals, intentions and criteria for success is a basis for the value of self-assessments and evaluations. He also points out that there is compelling evidence on the impact of goals for enhancing performance and self-efficacy (p.163 -165).

To evaluate whether upskilling is effectively realised, **Kirkpatrick's (1975; 2006) taxonomy distinguishes between four levels**:

1. learners' **reaction** to upskilling experience such as satisfaction and engagement (e.g. Cunningham, 2007; So & Brush, 2008; Athiyam, 1997; Wang, 2003);
2. **learning** (i.e. change in attitude, knowledge and skills, e.g. Bloom, 1956; Oliver et al., 2004);
3. **behaviour** (i.e. changes in practice, e.g. Chauhan et al., 2016; Xiao, 1996);
4. **results** (i.e. targeted organisational outcomes such as fixing a perceived performance deficit).

Various review studies conclude how the levels of learners' reaction and learning are frequently addressed in effectiveness research, while evaluations of impact on the levels of behaviour and results are limited (Berndt et al., 2017; Bluestone et al., 2013; Cant & Levett-Jones, 2021). A frequently researched conceptualization of behavioural effectiveness is 'transfer of training'. Transfer of training can be defined as the application, generalizability, and maintenance of newly acquired knowledge and skills (Ford & Weissben, 1997; Cheng & Hampson, 2008). See Appendix 7.1.1 for an overview over predictors of training transfer.

The effectiveness of upskilling interventions can be assessed objectively based on knowledge tests and observation, for example, **or subjectively based on self-evaluation**. Self-evaluation can be defined as the individual's judgement or appraisal of their own work, usually against a known or defined set of criteria (Rolheiser & Ross, 2001), whereas self-assessment depicts the ongoing introspective practice of self-reflection that prompts individuals to their behaviours, modulating and regulating their experiences (Pisklakov, 2014). Nevertheless, there seems to be a consensus in literature on the positive effects of both self-evaluations and self-assessments. Studies have found that self-evaluations potentially impact learners' performance, self-efficacy and intrinsic motivation (Rolheiser & Ross, 2001). Self-assessment has been found to increase certainty of self-knowledge (Sedikides & Strube, 1997), and of presenting itself as a learning opportunity (Pisklakov, 2014).

The reliability of self-evaluations has long been debated, with inconclusive results. In a meta-analysis on the validity of self-reported scores, Kuncel et al. (2005) revealed that the reliability of the scores increased at higher educational levels, implying that more advanced or educated learners dispose of more tools to perform such assessments. However, individuals may also be overconfident in relation to their performance in learning, and attribute higher self-scores

(Rogaten & Rienties, 2021; van Uum & Pepin, 2022). In a recent review of research on self-assessment, Andrade (2019) has pointed out that literature sustains a positive association between self-assessment and learning outcomes. Rolheiser (1996) posits that self-evaluation (which encompasses self-assessments) affects self-confidence. A positive impact on self-confidence may lead learners' setting higher goals and committing more effort to learning, which may yield higher achievement. The same phenomenon was noted by Pisklakov (2014) in a study on the role of self-evaluation and self-assessment in medical education.

Workplace learning approach

From the workplace approach to learning, research is not primarily aimed at demonstrating the effectiveness of predefined formal upskilling interventions. Research in this domain is concerned with **understanding how learning takes place and how it can be generated and enhanced in the workplace**. Tynjälä's (2013) 3-P model, which is based on the original 3-P Model of Learning by Biggs (1999), describes workplace learning by three interrelated components, including a presage, process, and product component. The presage component describes learning factors and the relationship between learner factors and the learning context. The process component concerns the nature of learning processes, describing work processes through which learning takes place, such as doing the job itself, reflecting on one's work and participating in networks. Finally, the product component in workplace learning reflects diverse learning outcomes such as task performance or the development of new roles.

Regarding research on workplace learning, three metaphors offer guidance in understanding and analysing learning (Sfard, 1998; Paavola & Hakkarainen, 2005; Ruiz-Calleja et al., 2017).

1. Closely linked to the conventional approach on learning, the **knowledge-acquisition (Sfard, 1998) metaphor conceives learning as an individual process of constructing and representing pre-existing knowledge**. Research representing this metaphor treats individuals as the basic unit of analyses and is mainly concerned with the individualistic process that leads to the possession and transmission of knowledge. Ruiz-Calleja et al. (2017) identify self-regulated learning, reflective learning, problem-based learning, and scaffolding as central educational concepts in this metaphor.
2. **Learning can be also viewed as a matter of social enculturation and the process of mastering the community's knowledge (*participation metaphor*;** Sfard, 1998). This metaphor assumes that learning happens by participating in collaborative practices. Through the contextualised and activity-based socialisation, learners adopt the skills that are recognized by the professional community and develop themselves from a novice to an expert. Learning communities, communities of practice, situated learning, and social constructivism are relevant educational concepts (Ruiz-Calleja et al. (2017).
3. The third conception is the ***knowledge creation metaphor* (Paavola & Hakkarainen, 2005) which conceives learning as the collaborative and systematic development of new knowledge** through common objects of activity (e.g. materials, concepts, and technologies). Ruiz-Calleja et al. (2017) see knowledge building, organisational knowledge creation, and meaning making as central educational concepts.

Concluding, if we want to evaluate upskilling interventions that follow a workplace learning approach, it is important to take a broad perspective and not only look at individual knowledge acquisition, but also successful socialisation and joint knowledge creation.

3 PROVIDING GUIDANCE FOR EVALUATING UPSKILLING INTERVENTIONS

Based on the understanding of the status quo of evaluating upskilling interventions in practice and literature, we **co-designed a step-by-step guidance with the BUSLeague consortium in T5.3. to evaluate use and outcomes of upskilling interventions, called the "EVALUATION123"**. It consists of an evaluation framework suitable for a broad variety of interventions and a toolset to support the self-directed application of the framework. Aside from informing about evaluation of effectiveness and providing respective guidance, the EVALUATION123 raises awareness about the benefit and use of learning analytics as well as learning paths. These represent the next steps to further improve interventions and increase the motivation of the workforce to engage in upskilling interventions.

3.1 INITIAL EVALUATION FRAMEWORK AND RUBRIC

The initial BUSLeague evaluation framework was established after the analysis and exploration phase of the design research and initiated the design and construction phase. The evaluation framework is based on a simplification of the Kirkpatrick (1975; 2006) model for training evaluation.

The **proposed framework focused on the first three levels of Kirkpatrick** (1-Satisfaction, 2-Learning, 3-Behavior), where evaluations can be designed and supported. The fourth level (4-Results) was excluded as this type of information is confidential and not accessible to external parties such as the BUSLeague stakeholders. The **levels of evaluation were named Level 1 – Reaction to Training, Level 2 – Learning from Training; and Level 3 – Change in Behaviour due to Training**. While the first two levels aim at measuring satisfaction and effectiveness directly after an upskilling intervention, level 3 measures effectiveness one to three months later in terms of its impact on practice.

Items on level 1 of educational effectiveness of training to be scored by TRAINEES								
1.1	Satisfaction with training (Cunningham, 2007, slightly adapted to better match target group)	Not satisfying at all	Not satisfying	Neutral	Satisfying	Very satisfying	Not relevant	Unknown
	Mastery experiences							
	The opportunity to learn new skills							
	The degree to which I improved on particular skills							
	How much I learned about how to perform better in this activity							
	My improvement in performance							
	My opportunity to practice new skills							
	Cognitive development							
	What I learned concerning the technical aspects of the activity							
	How much I learned about the various strategies used in performing the activity							
	What I learned about the basic content of the activity							
	The knowledge about the fundamentals of the activity I have gained							
	The extent to which I learned the essential concepts of the activity							
	Teaching							
	The quality of the overall instruction							
	The clarity of the instructor's lessons							
	The instructor's enthusiasm during the training							
	The empathy the instructor showed for the students in the training							
	The instructor's ability to effectively communicate content matter							
	My performance compared to others in the training							
	Normative success							
	The superiority of my skills in comparison to others in the training							
	How I am able to perform better than other students in the training							
	My skills compared to others in the training							
	My ability to outperform others in the training							

Fig 1: Excerpt from the rubric for the evaluation framework: level 1 (1st part of scales on satisfaction).

The **framework suggests the implementation of both subjective and objective measurements** for evaluation upskilling interventions, which is in line with the recommendations of the European Qualifications Framework. Whereas Kirkpatrick's model is more flexible and adopts experimental survey tools and assessment techniques, the proposed framework **adopted validated instruments for measuring the first three levels and design principles** for training programs that were identified in the literature study. As well, it specifies who uses ("scores") the instruments, the trainer, training designer or trainee.

Items on level 2 of educational effectiveness of training to be scored by DESIGNERS OF TRAINING OR TRAINERS								
2.1	How it is assessed	Not applicable at all	Not applicable	Neutral	Applicable	Strongly applicable	Not relevant	Unknown
	Trainees are assessed by (pre and post) knowledge testing (performed by trainees themselves)							
	Trainees are assessed by (pre and post) skills observation (performed by trainer, manager or researcher).							
2.2	Level of learning (Bloom (Oliver et al. 2004; Bloom)	Not applicable at all	Not applicable	Neutral	Applicable	Strongly applicable	Not relevant	Unknown
	After the training, trainees are able to remember (i.e. recall facts from the training material).							
	After the training, trainees are able to comprehend (i.e. understand, translate, and interpret the training material material).							
	After the training, trainees are able to apply (i.e. use the knowledge from the training material in a new context).							
	After the training, trainees are able to analyse (i.e. identify relationships).							
	After the training, trainees are able to synthesize (i.e. (re)assemble parts into a new whole).							
	After the training, trainees are able to evaluate (i.e. make judgements).							

Items on level 3 of the educational effectiveness of training to be scored by DESIGNERS OF TRAINING OR TRAINERS 1 month/3 months/1 year after training								
3.1	How it is assessed	Not applicable at all	Not applicable	Neutral	Applicable	Strongly applicable	Not relevant	Unknown
	The trainees' transfer of skills (depending on the topic of the training) to the workplace is assed by *observations* in the workplace (performed by trainer, manager or researcher).							
	The trainees' transfer of skills (depending on the topic of the training) to the workplace is assed one month after the training by *log/ user data* in the workplace (that is automatically generated).							
Items on level 3 of educational effectiveness of training to be scored by TRAINEES 1 month/3 months/1 year after training								
3.1	Self-perceptions (Chauhan et al., 2016)	Not applicable at all	Not applicable	Neutral	Applicable	Strongly applicable	Not relevant	Unknown
	Usage of skills acquired from training has helped me improve my work							
	I can complete my work faster than I could before attending the training							
	I can complete my work in a better way after attending the training							
	The quality of my work has improved after using skills acquired from training							
	I make fewer mistakes in job when I use the skills I have acquired from training							

Fig 2: Excerpt from the rubric for the evaluation framework: level 2 & 3.

Therefore, a **rubric was created for the evaluation framework**. For Level 1 (reaction to training; see Figure 1), the rubric includes instruments indicating the trainee's satisfaction in terms of perceived utility and relevance of the training. At level 2 (learning from training, see Figure 2), the rubric suggests the use of a knowledge test for trainees, and training designer or trainers' observation and assessment of learning goals with Bloom's revised taxonomy (Oliver et al., 2004). The **learning goals taxonomy includes six "levels of learning", which attribute the learner the competence to handle increasingly complex educational objectives**: i.e., remember, comprehend, apply, analyse, synthesise & evaluate. At level 3 (change in behaviour due to training, see Figure 3), the rubric suggests the addition of a trainee's self-perception survey to the observational assessment through the trainer or training designer, based on the work of Chauhan et al. (2016).

Items on the design of trainings to be scored by DESIGNERS OF TRAINING OR TRAINERS							
0	Design principles (Saks & Belcourt, 2006)	Not applicable at all	Not applicable	Neutral	Applicable	Strongly applicable	Not relevant	Unknown
	Pretraining activities							
	Trainee input and involvement							
	Employees are given advanced notification about training content prior to attending a training program.							
	Training needs of employees are identified prior to training.							
	Employees have precourse discussions with their supervisors prior to attending a training program.							
	Employees have input in decisions about training program content and/or methods.							
	Supervisor involvement							
	Employees are given release time to prepare for a training program.							
	Supervisors discuss the content and benefits of a training program with employees prior to a training program.							
	Supervisors participate in advance orientation or training sessions regarding the training programs to which they will send their employees.							
	Supervisors set goals with employees that focus on improving specific skills before employees attend training programs.							
	Training attendance policy							
	Employees have a choice as to whether or not they will attend any particular training program.							
	Attendance at training programs is voluntarily							
	Employees from the same department or functional group are trained together.							
	Trainee preparation							
	Trainees are given preparatory reading prior to attending a training program.							
	Training programs include activities or assignments that trainees are required to do before they arrive for the actual training program.							

Fig 3: Excerpt from the rubric for the evaluation framework: design principles (part on pretraining).

The rubric also comprised a scale for self-assessment of training designer or trainers' adherence to **design principles for evaluations** (see Figure 3). The latter is based on Saks and Belcourt's (2006) work that identified the influential variables for transfer of training in organisations.

3.2 CO-DESIGN FOR ACHIEVING AN UNDERSTANDABLE AND APPLICABLE GUIDANCE

The evaluation framework was first introduced to the BUSLeague consortium at the fourth consortium meeting, together with the rubric and entailed instruments. The partners got the

chance to answer the rubric for their upskilling interventions and discuss its value and shortcomings as well as needs in three subgroups. The results are summarised in Table 3.

The framework was seen to be insightful and useful for guiding training and evaluation design, but the rubric alone was perceived as complex to understand and insufficient to respond to the practical needs in evaluating the effectiveness of their upskilling interventions. One participant proposed that a "paper format, hand book or so will work! And then with a QR code that directly goes to a website etc." for accessing further tools.

Value	Shortcomings & Needs
• Design principles helpful for designing the training and checking its quality (also by other bodies/ stakeholders): "functions as a curriculum for how to design trainings" • Interest in application of the validated instruments • Appreciation of scales on L1 – Reaction to Training, especially "teaching", "interaction with others" and "fun and enjoyment" • Confirmation of importance of "Levels of learning" (L2 – Learning from training), esp. comprehend, analysis and evaluate • Agreement with the need of pre and post-test for measurements	• Absence of explanation of evaluation framework underlying rubric, and rubric itself including scales, items and their relevance • Tools are missing to ease the application of the evaluation framework and rubric • Consider level of education and literacy of trainees (understanding of items – mind language and concepts) • Scale of L1 – Reaction to Training: "Cognitive development" scale perceived ambivalently • Rubric is extensive – mind that evaluations must be focussed and short • Consider conditions of training: voluntary vs. mandatory, off the job vs on the job etc.

Table 3: Feedback on the evaluation framework and respective rubric.

3.2.1 Activities and approaches of the co-design trajectory

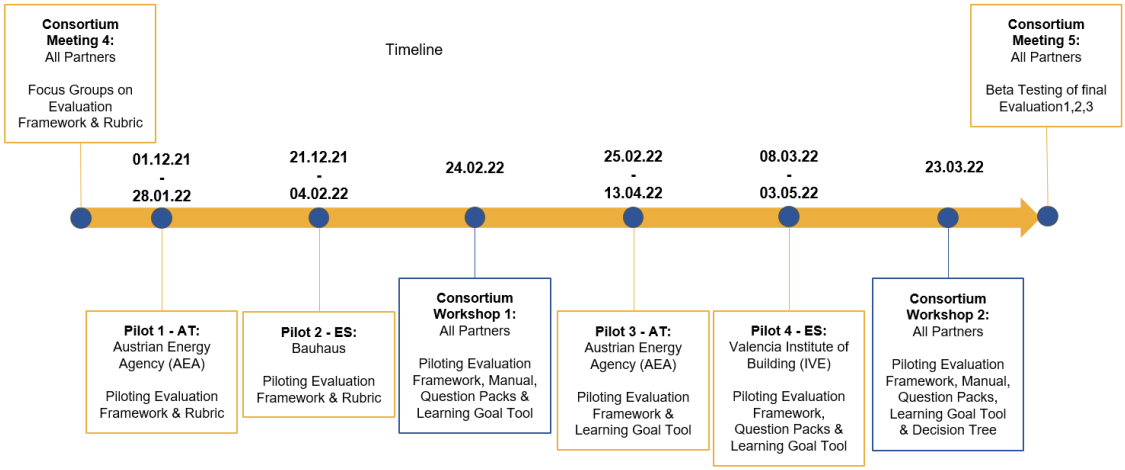


Fig 4: Co-Design Activities with BUSLeague Consortium for EVALUATION123.

The feedback from the fourth consortium meeting indicated that partners need additional explanations and tools to autonomously design and implement effective evaluations for their interventions. Therefore, we decided to **develop guidance, the "EVALUATION123" (see Section 3.4 for final EVALUATION123)** in collaboration with the BUSLeague consortium. This co-design trajectory should also help understanding what factors could prevent the adoption of the framework or the rubric and what could be done to address them.

The respective **design and construction phase of design research trajectory included six co-design activities** (see Figure 4) in the form of pilots or consortium workshops. The pilots constituted self-directed applications of the EVALUATION123, which were assisted and feedbacked by the University of Twente. Each pilot followed the structure:

- *Briefing*: Clarification of evaluation needs, goals and opportunities, and introduction to evaluation tools
- *Implementation*: (Observation of) Self-directed evaluation process and assistance in troubleshooting issues
- *Debriefing*: Discussion of evaluation results and interpretation, as well as usefulness and feasibility of evaluation process and tools (Semi-structured interview)

The consortium workshops adopted the same structure but in a more succinct form, focussing on the feedback of the complete consortium and promoting shared decisions of the partners. To give room for discussions on important topics, each workshop incorporated focus groups with some leading questions.

3.2.2 Main conclusions from the co-design trajectory for the EVALUATION123

We concluded from the six co-design activities that the **EVALUATION123 must provide a step-by-step guidance for understanding the evaluation framework and rubric**. The guidance needs to be simple, usable and valuable in terms of predefined evaluation goals so that it can be implemented autonomously even by inexperienced evaluators. Otherwise scarcity of time, human and material resources restricts its adoption in practice.

The **one-and-a-half-year co-design trajectory** of the EVALUATION123 **elaborated on design requirements and propositions and developed four design artefacts (see Appendix 7.2.3)**, which can be seen as the requested 'handbook' with an auxiliary toolset:

1) **Reference Guide (Manual):** The digital handbook explains the evaluation framework and central concepts to design an evaluation - quickly consultable and referable. Purpose and functioning of an evaluation are depicted in simplified visual way.
2) **Learning Goals Tool (Google Spreadsheet):** The tools assists the definition of learning goals to produce clear, viable and measurable training objectives following Bloom's systematic by providing partial structure, formulations and examples.
3) **Question Packs (Google Forms)**: The packs separate the instruments of the rubric into meaningful sets that can be easily replicated and combined into new surveys including automatically produced descriptive statistics.
4) **Decision Tree (Graphic):** The visualized process promotes design decisions evaluation on the three evaluation levels by triggering reflection of benefits (intended evaluation goals) and efforts (human, financial and material resources). It supports the selective reading of the reference guide in dependence of the evaluation goal.

Our application of Kirkpatrick's evaluation model in the **EVALUATION123 involves subjective and objective measures on all three levels of evaluation** (see Table 4). Whilst the evaluation of design principles can be subjective or objective (self-assessment vs. observation), learning goals provide subjective measures on level two and three.

		Kirkpatrick's model	EVALUATION123
Level 1	objective/ subjective		*design principles (question packs)*
	subjective	reaction	*reaction (question packs)*
Level 2	objective	knowledge/ skill test	*knowledge/ skill test (guidance)*
	subjective		*learning goals (learning goal tool)*
Level 3	objective	knowledge/ skill observation	*knowledge / skill observation (guidance)*
	subjective		*learning goals (learning goal tool)/ motivation (question packs)*

Table 4: Clarification of the subjective and objective measures of the EVALUATION123.

3.3 BETA TESTING OF "EVALUATION123" AT FIFTH CONSORTIUM MEETING

On the fifth consortium meeting on the September 21st 2022, we conducted a **final evaluation of the EVALUATION123 reference guide and toolset** that focussed on the support of evaluation design. 14 participants of the 9 partners of the BUSLeague participated in the workshop: IGBC (IE), TUS (IE), IVE (ES), BAUHAUS (ES), AVE (FR), ISSO (NL), AEA (AT), IRI-UL (SL) & BCC (BG). Several partners were about to conduct a training and in the need of an evaluation plan. Therefore, we adopted the structure of the second consortium workshop in expanded form. The activity provided more time, allowed the usage of the own upcoming training as case and increased the interaction with the reference manual and toolset.

Whilst the most prominent evaluation plans and respective evaluation contexts are used to demonstrate the EVALUATION123 in the subsequent chapter, we report the results of a survey and plus-minus test (+/-) in this section. Results of the beta-testing were leveraged to improve the EVALUATION123. We present the resulting final versions in the next section.

3.3.1 Understanding usability and usefulness of the EVALUATION123 - Survey research

We explored the **usability and usefulness of the EVALUATION123** with a questionnaire. The partners indicated their agreement with statements about the reference guide, decision tree, learning goals tool and question packs (see Appendix 7.3):

- **Usability of the reference guide:** The partners agreed that it is easy to look up things and that the images and layout facilitate comprehension. They only somewhat agreed that its text is easy to read and uses too many technical words.
- **Usefulness of the reference guide:** The partners agreed that the reference guide helps to understand evaluation needs and possibilities, provides a good overview over the different possibilities and is useful for planning evaluations for different contexts. They also agreed to feel able to plan an evaluation using the reference guide.
- **Usefulness of the decision tree:** The partners agreed that it helps planning an evaluation, and helps getting the maximal evaluation outcomes within their possibilities (ad. trade-off evaluation benefits and efforts).
- **Usefulness of the learning goals tool:** The partners agreed that it helps in designing their evaluations. They somewhat agreed to the learning goals tool helping in specifying the ULOs and clarifying learning objectives.
- **Usefulness of the question packs:** The partners agreed that the packs help in selecting items for their survey. They strongly agreed the digital format in Google Forms facilitates their use.

The results indicate that the BUSLeague consortium, represented by all but one partner, perceived the reference guide as usable and saw the reference guide, decision tree and learning goals tool, question packs as useful for conducting evaluations of their trainings. They did not disagree with any of the statements about usability and usefulness. Yet, there is still potential to improve the bridge of the learning goals tool to existing ULOs and increase the comprehensibility of the reference guide. For the latter, we draw upon the results of the plus-minus test (+/-).

3.3.2 Improving on EVALUATION123 reference guide – Plus-Minus Test

We **qualitatively explored the comprehensibility of EVALUATION123** with a plus-minus test (+/-). The partners were asked to mark and comment sections they used alongside the design of their evaluation plan: i.e., they should indicate their appreciation with a plus (+) or (-), and provide additional reasoning for their rating. In total, they rated 62 sections as positive and 22 as negative. In the following, some meaningful insights will be summarised.

The introduction to evaluations and the paragraph on "Asking the right questions" were appraised: "I like the simple, straightforward language". This means that the style and language seem to be appropriate and not overly complex or technical. This also applies to the explanation of the goals and aims of evaluations that appeared reasonable for representatives of the construction and building domain and in line with their expectations. In general, figures were appreciated.

As well, the partners perceived the section "Reflection on availability of resources for evaluation" as applicable. This confirms that the identified limitations of time, human and material resources are relevant in the construction industry. The decision tree and respective reflection prompts received many positive markings, except for one user seeing these considerations as not relevant. It was suggested to mirror the graphic of the decision tree for ease of reading and highlight the starting point.

The presentation of the rubric in form of the questionnaire packs as well as items got mixed reactions. Comments referred to "vague and unclear definitions". The explanation of the learning goals via the SMART model (Doran, 1981) and the learning goals tool was very positively perceived. The understanding of the learning goals tool seemed to have been consolidated and valued by users at this time of the project. They also started seeing learning goals as the focal point for both training design and evaluation efforts.

In addition, the evaluation on Level 3 was in the focus and received several comments. This shows the partners' focus on transfer of learning, despite not necessarily being able to assure all preceding steps. They are interested in the workplace learning approach to upskilling, but understand the respective transfer of training as process complex to realise.

Finally, learning analytics and learning paths were appraised. Yet, comments suggested that their benefits need to be clarified further. Appendices were overall rated positively, which mostly reproduced the rubric and question packs. This indicates again that the partners see the scales as valid, but opt for measurements with only a few items (/questions) in practice.

3.4 TEASING "THE EVALUATION123" FRAMEWORK AND TOOLS

The EVALUATION123 guidance resulted from the collaborative design effort of the BUSLeague consortium, considering all partners' problems and needs. In the following, we **tease the final Evaluation123 guidance with all of its elements**. This means that we present excerpts from the reference guide (see Figure 5, see Section 3.4.1) on:

- the learning goals tool (see Section 3.4.2),
- the questionnaire packs see Section 3.4.3) &
- the decision tree (see Section 3.4.3).

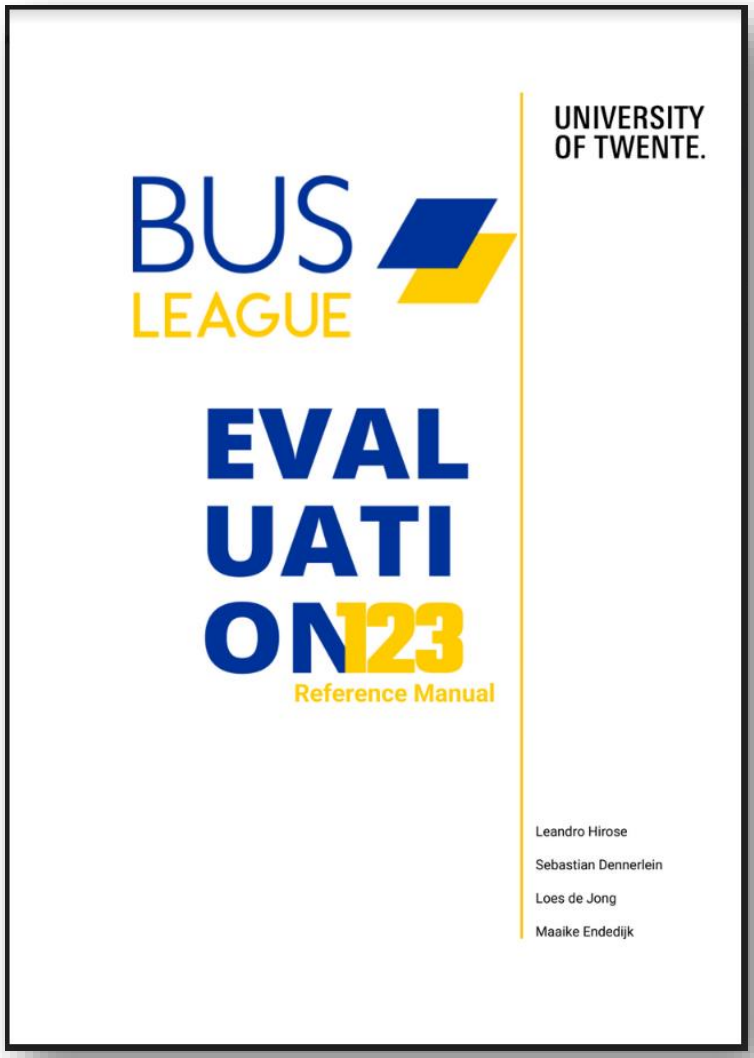


Fig 5: Front cover of Reference Guide of the EVALUATION123 and its Learning Goal Tool.

The **explanations can only be illustrated as the explanations of the reference guide are too comprehensive and the auxiliary tools are web tools**. The full reverence guide and toolset can be accessed via the following Google Drive Link and copied for use and adaptation: https://drive.google.com/drive/folders/1sKn0ZBdj__neC8BdCpqW_d9NcabfCrM3.

3.4.1 Reference Guide – Explanatory Manual (Snippet from Introduction)

Designing evaluation programs may seem a fairly straightforward process for the unwary evaluator, and a complete nightmare for a diligent, but unexperienced one. The aim of this reference manual is to provide guidance on how to design an evaluation program. You will also be introduced to a number of supporting tools that will help you designing and implementing your evaluations.

Why evaluate?

At first glance, this may seem a vague, or even naïve question. We evaluate because we want to know – may be a common response to this question. But our focus should be a little further, and we should ask ourselves: for which purpose are we evaluating?

The answer to this question is more varied and clarifies the true nature of the purpose of evaluating. Do we want to assess knowledge and skills of trainees? Do we want to measure the effectiveness or impact of our training? Do we want to improve our training? Do we want to use these results to award certificates of qualification? Different goals may require different approaches. For example, if your intention is to certify trainees, you will need objective knowledge and skills tests, whereas in simpler micro trainings you may be satisfied with more subjective measurements of achievement. Having these purposes clear from the beginning is essential to keep our evaluations true to what they intend to measure.

Asking the right questions

> ***"Good questions inform, great questions transform"***

Posing the right question is a difficult art. An excessively precise question may be informative, but it does not allow us to draw general conclusions on the training. On the other hand, questions that are too broad or vague may not provide the specific insights we are looking for. So, a good approach to this problem is to respect important perspectives on training in our questions, so that the combination provides us with the information we need. For example, we may ask participants to rate the quality of the training they just received. The answer will only inform us about the quality of what was delivered, but we cannot know from this whether they will apply this new knowledge to their work. However, we can additionally ask them to rate their perception on the relevance of the topics discussed during the training in their working practices. The different combinations of both answers provide different insights. A high rate in both questions may indicate that the training was successful and that the participants are willing to make use of the new knowledge. A high rate on the quality but low on relevance may indicate that the participants did not understand the relation between what was taught and its utility. This might be due to a flaw in the design of the training (too difficult or unrelatable) or in the communication of the learning goals. Conversely, a low rate on quality but high on perceived relevance may show that participants felt that the content was insufficiently approached or too superficial. Clarifying on the remaining questions requires asking the right questions. This reference manual provides a structured overview on important perspectives on training evaluation, and proposes questions based on standardized and validated instruments. The choice of proper questionnaires is very relevant, as the formulation and selection of meaningful clusters of questions is an art or discipline in itself.

Structuring an evaluation

The EVALUATION123 was developed to facilitate the design of evaluation programs. Based on state-of-the-art evaluation approaches and methods, it will help you get the insights in your trainings that you are interested in. It consists of three different steps (see Figure 6), adapted from Kirkpatrick's training evaluation model. At level 1, you will assess the quality of our

programs by measuring the strength of the design principles, and the satisfaction of trainees with regards to the quality and utility of the training. At level 2 you will measure how much they have learned and their perception of their progress. This tells us whether our training is effective and how likely the new knowledge will be transferred to work activities. A few weeks after the training, you can assess level 3 – it tells us the impact our training had on their professional practices. The classic Kirkpatrick model has a fourth level – Results – usually measured as the financial or organizational impact the training created at a business level. Because this measurement relies on internal company data and is often not accessible to training and/or evaluation providers, the fourth level was not included in the EVALUATION123.

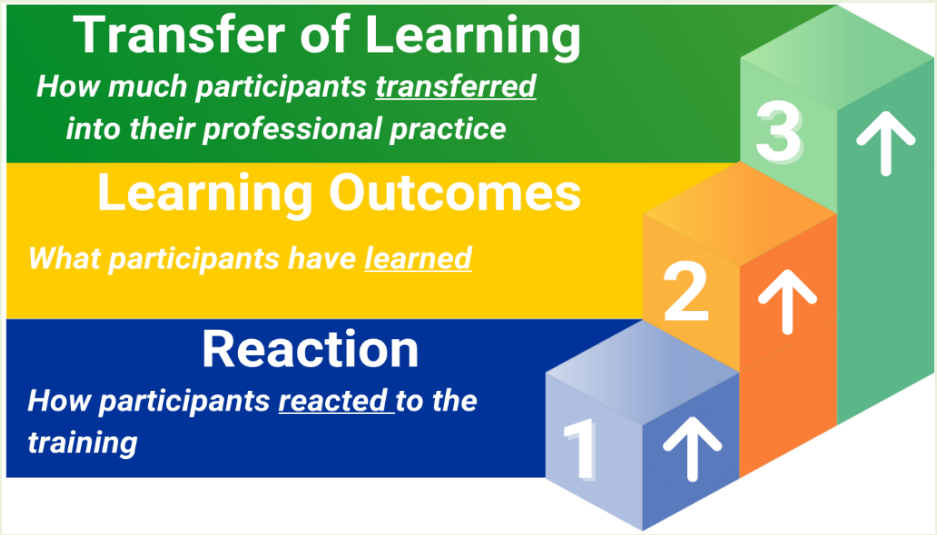


Fig 6: Evaluation123 3-level model

What do I need to implement an evaluation?

Evaluations, just like any other activity, demand resources. These can be **human** (people who will be involved and coordinate the process), **material** (technologies, documents, forms, spaces, budget), and **time** (from both evaluators and evaluands). Several factors can influence your potential for setting up an evaluation program. Clarifying what resources are available to you will help you mould and adapt the process and make the most out of them.

Human Resources

Not every organization has a professional evaluator or even a dedicated training department. In such cases, it is important to determine who will be able to conduct the evaluation process, activities and deliver the results. Different people can do these activities, but they require some coordination and planning. In order to determine your availability of human resources, try to answer these questions:

- *Who will manage and conduct the evaluation processes and activities?*
- *Who will collect and who will analyse the data? (administer surveys, correct and grade tests, produce reports)*
- *Who will ensure compliance with General data Protection Regulation[4] (GDPR) and ethics?*
- *Who are the trainees? Are there any particularities in that audience that I must consider?*
- *How will I be able to reach trainees for testing and surveys?*

Material Resources

Material resources refer to the tools that you will need to implement an evaluation. These can be hard (such as paper surveys and exams) or soft (digital) tools.

- *What means and technologies are available to me to set up an evaluation? (e.g., learning management system, learning analytics and statistics software)*
- *Which survey platform better suits my needs and requirements in terms of included automatic analysis, data storage, cost etc.? (e.g., Google Forms, Microsoft Forms, RedCap and Qualtrics - see EVALUATION123 – QUESTION PACKS below.)*
- *Where will data be stored?*

Time

Timing is not only essential for evaluating, but it is also an underlying factor influencing all the process. Evaluations require time from evaluators, trainees, and the organization.

- *How much time can I dedicate to the evaluation process?*
- *How much time can trainees dedicate to the evaluation process?*

3.4.2 Decision Tree

As you have seen, designing, and setting up an evaluation program requires a balance between what you desire and expect of it, and what you can afford to do. Determining this balance can be a bit easier when you are able to see how evaluation processes interact and what they require in terms of resources.

In EVALUATION123, this can be better visualized in the EVALUATION123 **Decision Tree** (figure 7). By answering a few direct questions, you will be able to determine the possible routes for your evaluation program, always taking in consideration your goals and capabilities. The decision tree starts with the **green box Evaluation123** at the top. **Yellow boxes** marked with a letter (A-G) are reflection and decision processes. The main question is represented in the box, but you need to refer to the corresponding explanations on the following page. They will also lead you to a deeper understanding and reflection on your abilities and limitations to reaching your evaluation goals. **Blue boxes will** inform about the processes triggered by each decision.

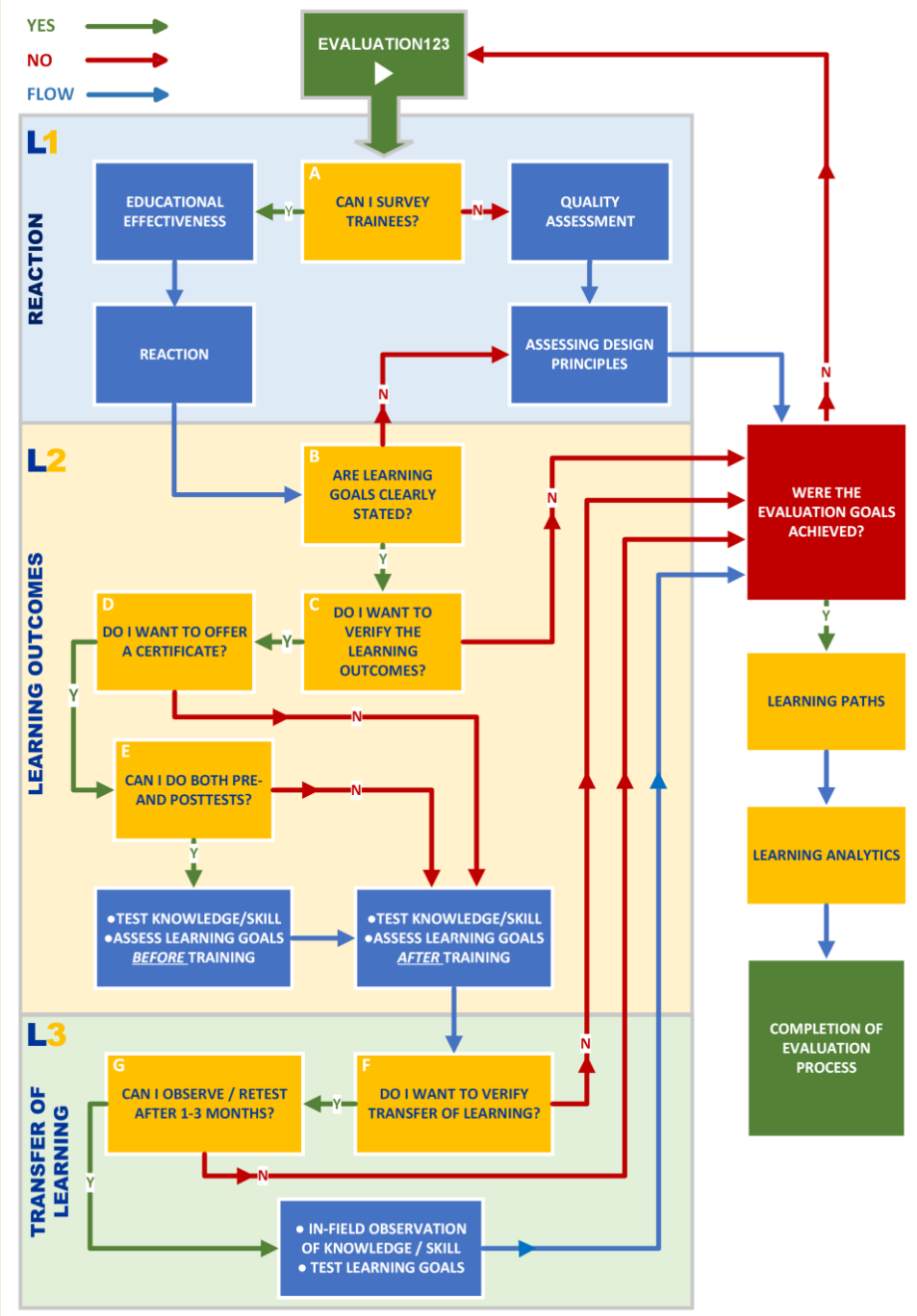


Fig 7: Evaluation123 - Decision Tree

*Letters (A-G) in the yellow boxes are **Guiding Questions** that will help you determine your capabilities and opportunities for reaching your evaluation [Next page].*

A - Can I survey trainees? (p.3)

- *Will I be able to reach trainees and collect data?*

Try to define what your opportunities for reaching the trainees are. You may want to check which medium are you going to use (online forms, e-learning platform, paper) or if you want to state the questions in interviews and analyse them qualitatively. Assessing the quality of an evaluation and design principles is always recommended, even when trainees cannot be approached. This corresponds to the Level 1 of the Evaluation123.

B - Are Learning Goals clearly stated? (p. 22)

- *Are my Learning Goals clearly defined for both trainer and trainees?*

Goals are the general aims or intended results of a program. Learning Goals can be derived from Units of Learning Outcomes (ULO).

C - Do I want to verify the Learning Outcomes? (p. 21)

- *Are the results on the performance of my trainees relevant for my training purposes?*
- *Can I dispose of the human, material and time resources required for an evaluation?*

This corresponds to the Level 2 of the Evaluation123 and will imply in administering tests to trainees. The complexity of the process will depend on the goal of the evaluation and intended results.

D - Do I want to offer certification? (p. 22)

- *Does my training aim at providing participants with certificates?*
- *Do I want my training to be recognized at a specific level?*

Trainings that aim for certification at European level follow the recommendations of the European Qualification Framework. Certificates may also be awarded at a local or internal level, in a simpler process.

E - Can I do a Pre- and Post-test? (p. 15)

- *Can I test trainees before and after the training?*

Testing both before and after the training increases the precision of evaluations, as it accounts for factors such as previous knowledge and sets a benchmark for comparison. In some circumstances, it might not be feasible to implement pre- and post-tests, but this is not an impediment for advancing in the evaluation.

F - Do I want to verify Transfer of Learning? (p. 35)

- *Do I want to measure the lasting impact of my training?*

This corresponds to the Level 3 of the Evaluation123. This level of evaluation demands more resources and time, but allows understanding the impact of learning on the daily work of the professionals. Those resources can be manageable if planned beforehand.

G - Can I observe / retest after 1-3 months? (p. 35)

- *Will I be able to implement an observation test to verify the impact of my training after 1-3 months? Can I still reach the trainees after 1-3 months?*

We can measure the lasting impact of a training by setting up observations of the trainees using their skills and knowledge on practice. If this is not feasible, we can still estimate this impact by reassessing Learning Goals and comparing them to previous results.

3.4.3 Learning Goal Tool

At level 2, you will examine the effectiveness of our training programs in terms of knowledge and skill acquisition and comparing them to the learning goals that were set for the program. Learning goals are determined before the design of the training and are also expressed as Units of Learning Outcomes (ULOs).

Units of Learning Outcomes - ULOs

The elaborating and definition ULOs are described in the document ***Database of common/harmonized descriptors***, and can be summarized as follows:

Selection of **Learning Goals**

Goals are the general aims or intended results of a program. They are broader concepts that may comprise one or more objectives.

Selection of **Learning Objectives**

Objectives are specific, measurable steps and tasks need to be accomplished to reach a goal. One or more objectives can direct to the same goal.

ULOs form a database of common and harmonized content for knowledge, skills, and competences, in the form of task-based qualifications, comparable across EU countries. For example, a ULO based on knowledge is to know about the different types of thermal insulation. A skill based ULO would be to determine the most cost-effective insulation for a building. Both are part of a competence, which is the ability to assess the energy demand and gains of a building. The European Qualifications Framework (EQF) makes the following distinctions:

Knowledge: result of the assimilation of theoretical and/or factual information related to a field of work or study.

Skill: the application of knowledge and ability or expertise to complete tasks and solve problems. They can be described as cognitive or practical skills.

Competences: the ability to apply knowledge and skills autonomously and with responsibility in a variety of contexts.

Self-assessment of Learning Goals

Why should I ask participants about learning goals? Can I trust a self-reported evaluation?

There are many reasons to include this practice in your surveys. The idea of having people self-grade their knowledge has been extensively studied[16–19]. While objective tests (see above knowledge and skill test) enable you to measure the concrete knowledge gained with the training, self-assessments of Learning Goals will tell you how the trainee perceived this increase. Also, the personal understanding of progressing or achieving a knowledge or skill is important, as the ability to properly assess oneself enables self-regulation[20] and self-responsibly determination the learning process.

Secondly, the assessment of learning goals helps to efficiently get an indication of training effects as compared to more costly formal tests (i.e., time, material and human resource) that require design and validation. For example, a formal test after a training on the efficacy of insulating windows would have to cover a variety of topics, such as types of glass, frames, opening systems, weather strips, shadings, etc. Such test for knowledge and skills with all the subtasks are extensive to objectively assess and they are often tedious for participants. To check whether the trainees feel they have learned the intended knowledge or skill, learning

goals can be leveraged to obtain high level understanding of the training effects. The subjective test is less accurate but can provide quick overview over assumed learning outcomes, as compared to more precise objective test. The decision on the kind of test depends on the needs of the evaluation such as providing certificates versus improving trainings.

Assessing learning goals may therefore provide information on the quality of programs. An exceptionally low rating on a learning goal may indicate that the topics related to it were not sufficiently covered during the training, for example. This may point out to opportunities for correction or improvement. As they relate to the perceived self-efficacy of an individual, self-reported learning goals may also help assessing whether the course content was adequate to the needs of trainees. Ultimately, giving the chance of participants to think about learning goals improves subjective assessments and may prepare them for the learning experience (pre-training) and reflect on their progress and learning needs (post-training).

Designing (pre- and post-training) Learning Goals Self-Assessments

As it was previously mentioned, learning goals are a component of ULOs. In well-defined ULOs, learning goals are easily identifiable as the main goals of the training programs. They describe, in general terms, what a learner should be able to do after the training and comprise several learning objectives (subtasks).

However, there may be cases where learning goals are not immediately available or have not been clearly defined. Learning goals are at the core of a training program, so it is important to define what makes a solid learning goal and how to build them. You will learn more about it when you start using the Learning Goals Tool.

Determining Learning Goals

Learning goals are S.M.A.R.T

S.M.A.R.T (specific, measurable, achievable, relevant, and time-bound) is a well-known set of criteria introduced by Peter Drucker as an approach to defining consistent and coherent goals. Here, we will show you how they apply to determining learning goals for the upskilling training programs in BUSLeague.

Specific: learning goals should be clear and specific. Goals with 'gaps' in their statement raises uncertainty about what is to be achieved. An effective way to test whether a learning goal is specific is by answering the following questions:

- **When** are they expected to do it? (e.g., after the training, by the end of the year, etc.)
- **Who** is expected to achieve the goal? (e.g., trainees, craftsman, employees, etc.)
- **What** do they have to do to achieve it? (e.g., being able to know, do, analyse, etc.)

Furthermore, in some cases, it may also be worth specifying the conditions or circumstances those goals apply to:

- **How** are they expected to do it? (e.g., independently, under defined circumstances, etc.)

Example:

After the training, **WHEN**	learners **WHO**	will be able to calculate the energy efficiency coefficient of a building. **WHAT**

Measurable: learning goals should translate into concrete metrics and milestones as indicators of progress. This is especially relevant when learning goals are too broad or general and not directly quantifiable. In such cases, they have to be reformulated into more specific goals that provide concrete measurements as indicators of achievement. For example, a training program may state as a learning goal that "Trainees should be able to reflect about the impact of thermal installations in the energy efficiency of a building." But what do you mean by reflection? And can you easily measure it?

Achievable: learning goals should be sufficiently feasible for trainees to acquire such knowledge and skills in the specified conditions and amount of time. For example, a training consisting of a short video explanation on how to install solar panels cannot expect inexperienced trainees to be able to successfully perform installations of solar panels.

Relevant: reflecting about its relative importance will fine-tune a learning goal to what is meant to be achieved. Several questions can assist in determining its relevance:

- Does it fit properly with the overall goal of the program?
- Is it the appropriate time to set this goal?
- Will the necessary resources be available to achieve it?
- Will the content of the program be sufficient to reach it?

Time-bound: determining time boundaries is important to contextualize the objectives of an assessment. Most post-test assessments can use immediate time frames, such as after the training, or at the end of the session. Some may be middle-run aims, such as by the end of the year, or after 30 hours of practice.

Using the Learning Goal Tool

The Learning Goal Tool (LGT) first helps you to analyse and define the level of learning[13] one specific goal is aiming for:

Conceptual Knowledge: ability to remember and understand ideas and concepts.

E.g.: identify, describe, name elements.

Procedural Skill: ability to apply the knowledge in a specific task

E.g.: calculate, sketch plans, install elements.

Analytical Thinking: ability to make judgements and evaluations based on knowledge and experience

E.g.: evaluate, assess, correct, judge plans.

This chart is a simplification of the Taxonomy for Learning, Teaching and Assessing, a well-known classification of learning objectives. Therefore, the first thing you see in the **LGT** is the chart of levels of learning (see figure 8). it includes a brief description of the levels and examples of activities belonging to each level:

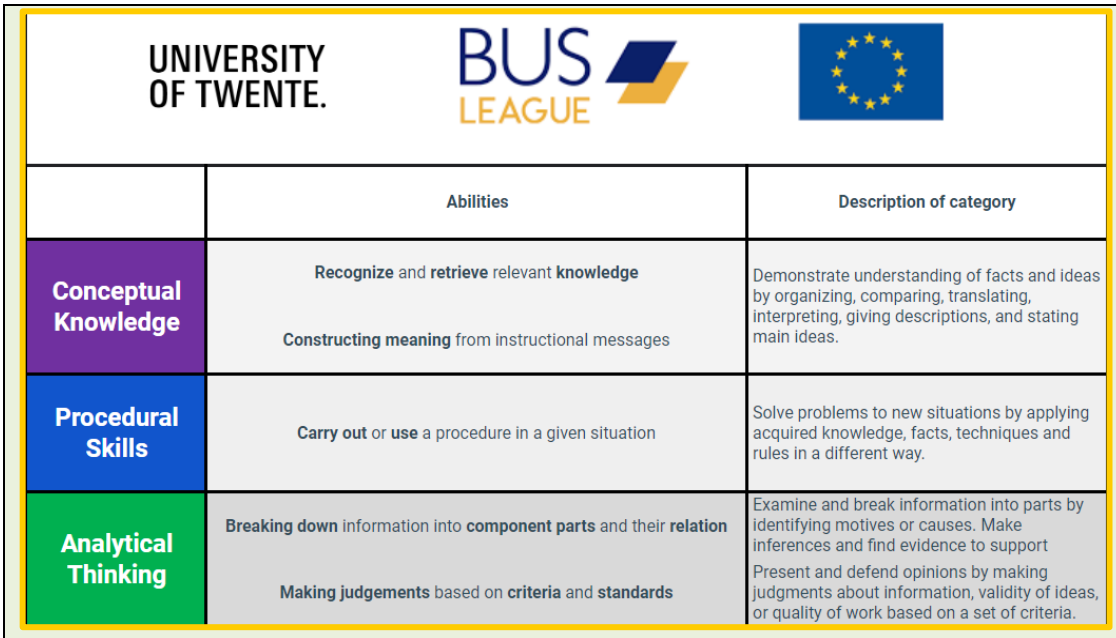


Fig 8: Interface of the Learning Goals Tool

The second feature you will see in the **LGT** are the boxes where you will specify the language, time condition and participants (figure 9):

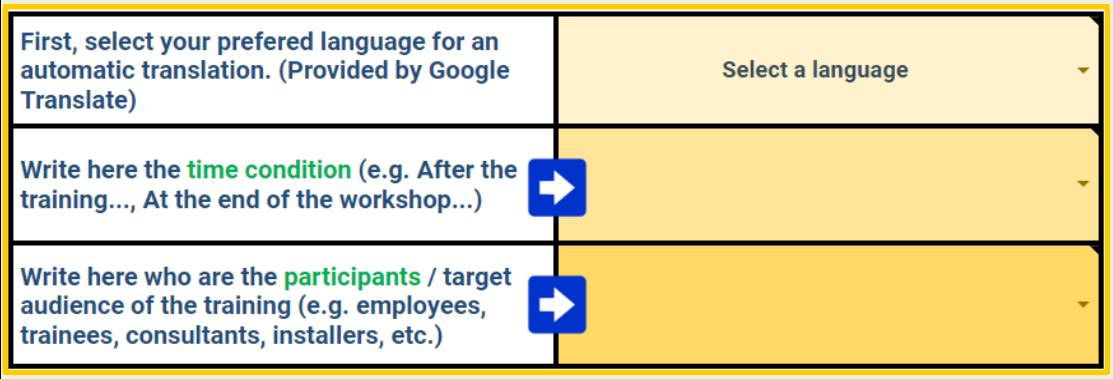


Fig 9: selection boxes of the LGT

First, an embedded translation tool is presented that will help you export the finished Learning Goals to the target language of the evaluation. This is an automated translation, so it is worth checking for errors, but simplifies the work.

Time condition refers to the time-bound aspect of a sound learning goal. More specifically, it refers to when you are performing the evaluation and are related to the pre- and post-training aspects of the design of an evaluation. Common examples are after the training, After the workshop, At the end of the course (post) or Before the training, Currently (pre). Alternatively, there is a preloaded drop-down list with a few suggestions you can use.

Participants refers to who is taking part in the training, the target audience: e.g., employees, workers, trainees or even customers. Once again, you will find a list with a few suggestions, but if you double click the cell, you will be able to create your own entry.

The next feature you will see in the **LGT** is the 'tool' for formulating the learning goals itself (figure 10):

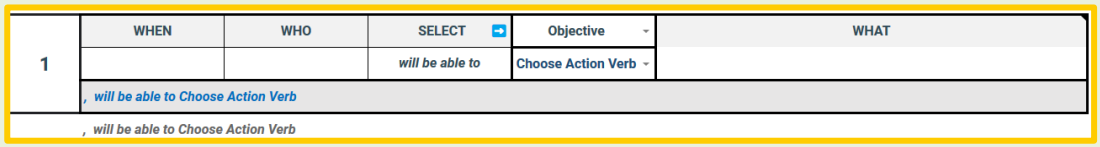


Fig 10: Learning Goals Tool

*__Note__: when you type down your options, you may see an '**Invalid**' warning. Just ignore it.

If you have filled the time-condition and participants box, they should appear here automatically.

Next step is to select the level of learning: conceptual knowledge, procedural skills, or analytical thinking. This will help you decide for the most suitable action verb. Each learning goal should have only **one action verb**, which may fit in one of the three levels. Because listing all possible verbs would be impracticable and make the tool difficult to use, we listed the most common verbs. However, we encourage you to reflect on the categories and add your own verbs. Is the goal of the training to gain knowledge? Acquire or improve skills? To be able to perform critical analysis? Suitable verbs are verbs that describe precisely what trainees are expected to do and that provide a clear indication for their measurement. For example, if you state in a learning goal that trainees should be able to **operate** and **explain** the functioning of an equipment, how do you validate a learning goal if they can operate but not explain it? In this case, those should be separate learning goals.

Lastly, it is time to fill in the **WHAT** box. Here you must describe the knowledge or skill that is part of the learning goal. As you can see, there is no verb in the **WHAT** section, and you should describe the executable elements that are part of the learning goal.

Here is an example:

WHEN After the training
WHO participants
Objective Conceptual Knowledge
Action Verb calculate
WHAT the energy efficiency coefficient of a building.

If you have completed all these steps, you should end up with a finished learning goal, like this (figure 11):

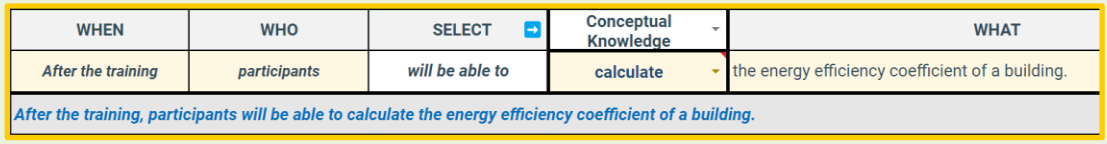


Fig 11: Outcomes of the Learning Goals Tool

Operationalization of Learning Goals assessments

Assessing learning goals is a straightforward process when they have been defined according to the above rules.

Post-training evaluation (post-test): executed **after** the training intervention.

Learning goals are usually assessed in terms of their achievement, which is often measured by percentage. This also allows to easily assess the progress between multiple points of measurement alongside the EVALUATION123. When asking participants to grade their knowledge of learning goals, you should always opt for the most specific **scale** available. If the tool you use for designing surveys allows you to use a 0-100 scale, go for that option. If you use another tool, such as Microsoft or the Google Forms in the EVALUATION123 packs, you may need to opt for the 1-10 scale.

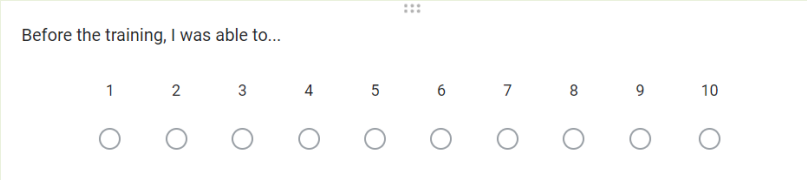


Fig 12: Example of a 0-10 scale

Besides, you need to differentiate two types of learning goals assessments:

Pre-training evaluation (pre-test): executed **before** the training intervention
Post-training evaluation (post-test): executed **after** the training intervention.

There are two possible scenarios for performing an evaluation. In **Scenario A**, you have a chance to do a pre-test, followed by the training and a post-test to evaluate the results. This is what we call a 'standard' scenario, and the one you should opt for whenever possible. However, there may be circumstances where you will only have one chance to evaluate your participants. In **Scenario B** the structure of the evaluation needs to be slightly adapted in order to gather the most information with only one test or survey.

3.4.4 Questionnaire Packs (Google Forms)

Throughout this reference guide, you will be presented with a collection of previously selected test and survey tools. As we have mentioned before, these instruments are standardized and validated, and have been slightly adapted to better suit your purposes.

The EVALUATION123 is accompanied by Question Packs. These packs contain all the proposed tools and surveys organized and identified for ease-of-use. When designing your evaluation, you can simply copy and paste the packs that are most relevant to accomplish your evaluation goals (more information on how to use the import function in Google Forms is found in the document Instructions for Using the Online Tools in the Evaluation123 Google Drive folder, or in Appendix A). You may also add to those packs any other questions that you find relevant for your organization. For example, you may want to learn more about the profile of your trainee and add some socio-demographic questions, such as age, education, or years of experience in the field.

In order to make the process more accessible to all, you will find them in Google Forms. The choice for a Google platform is due to its accessibility and familiarity to most users. Google Forms also allow you to import the Question Packs into your own surveys and tests, making it easy to organize and disseminate them. The results will be immediately available as a spreadsheet and it also offers some other useful features, such as (very basic) support for data analysis and visualization (which may be useful for inexperienced users) and can be easily converted and imported into other platforms, such as MSOffice or iWork.

Below you will find a quick description of the main options of survey tools currently available and their features. You may also check their GDPR compliance on the direct links provided.

Qualtrics – subscription based

- https://www.qualtrics.com/
  - https://www.qualtrics.com/support/survey-platform/getting-started/qualtrics-gdpr-compliance/

Qualtrics is a highly usable platform to design and implement surveys. It allows you to set up, share and collect the data all in one platform. Qualtrics offers a wide range of options for surveying and different scales and is capable of generating customized reports and improved visualizations. It also provides powerful and easy-to-use analytical tools for the interpretation of data and provides automatic reports.

[…]

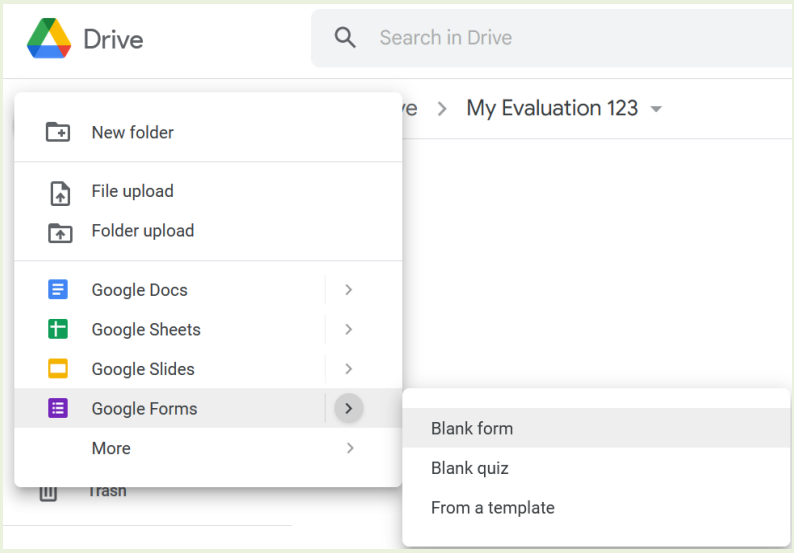
Google Forms - Free

- https://docs.google.com/forms/
  - *https://cloud.google.com/privacy/gdpr*

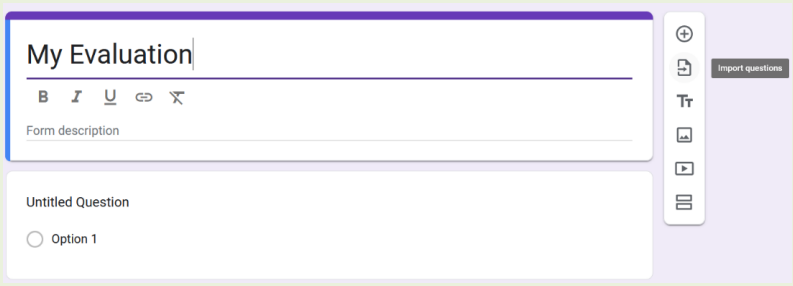
Google Forms is an integrated solution within their Drive options, which means it is totally compatible with other Google's solutions such as Docs or Sheets. It allows you to create and share surveys and analyse responses in real-time. It also exports raw data in spreadsheet format and generates a few automatic summaries and graphs.

IMPORTING QUESTION PACKS

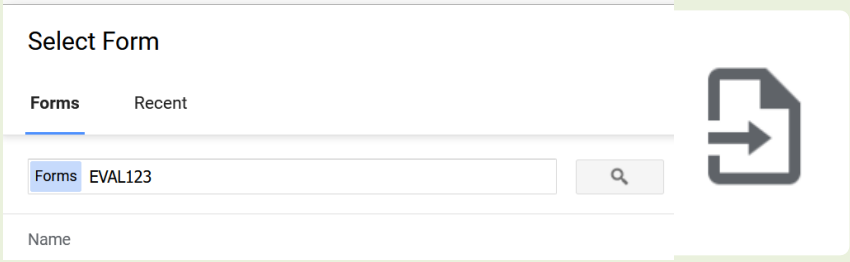
1. In your Google Drive folder, create a new blank form:



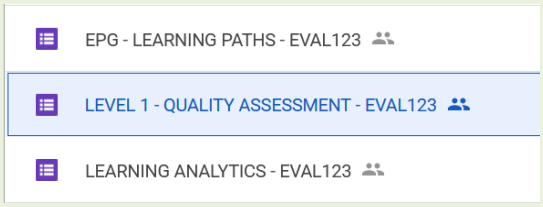
2. Select the Import questions function on the right-side bar:



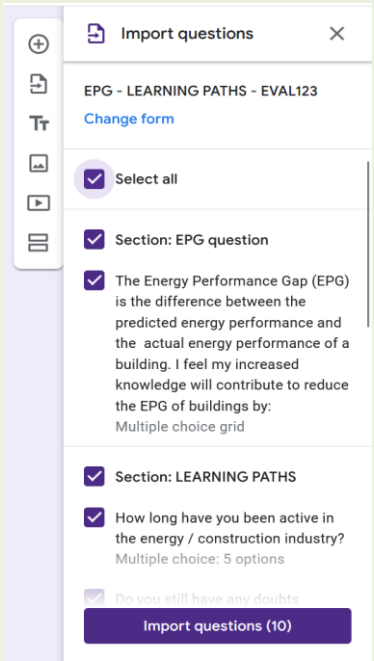
3. A Select Form prompt will open. Type in **EVAL123** to find all available packs.



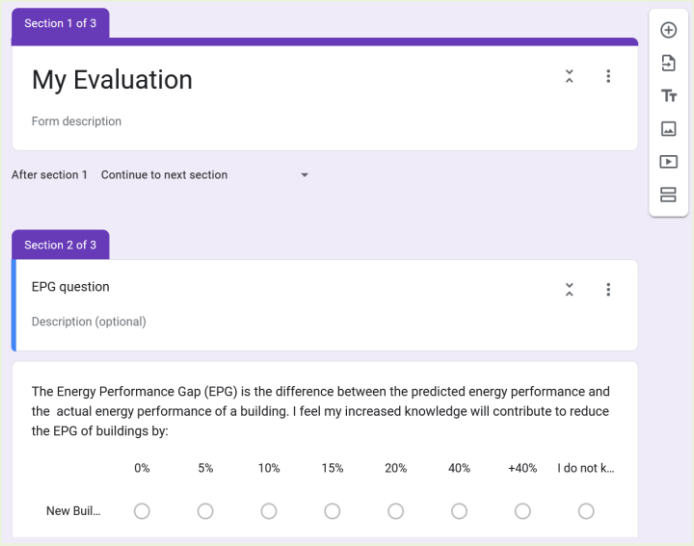
4. Select the packs you will use



5. A panel will open on the right side. Here you will be able to select questionnaire items.



6. You can repeat the process and import all other packs that will be part of your evaluation.



[See Figure 13 for an example of a Questionnaire and the Question Packs]

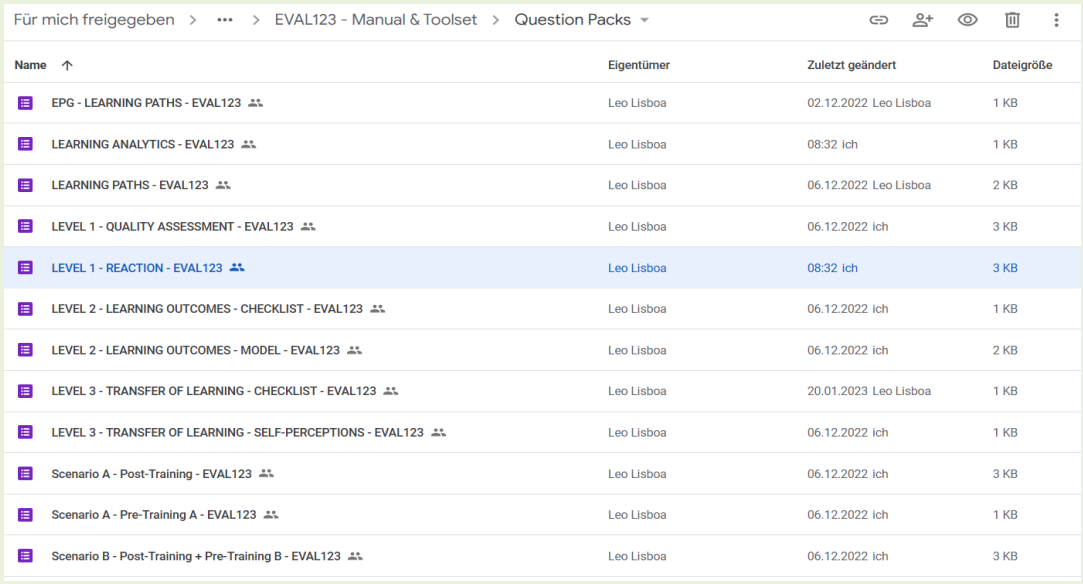


Fig 13: Overview over a questionnaire and question packs with for additions for learning analytics and paths.

3.5 LEARNING PATHS AND ANALYTICS TO FURTHER IMPROVE INTERVENTIONS

In the following, we will tease the last addition to the EVALUATION123 that shall **inform future evaluations and further improve interventions, namely the exploration of learning paths and learning analytics**. Both allow understanding motivational triggers and boundaries that professionals face when engaging into upskilling interventions. These opportunities have been too early for most partners of the BUSLeague consortium and could not be addressed within the runtime of the project. Therefore, we introduced the evaluation topics of learning paths and learning analytics and proposed some initial approaches to the consortium via the EVALUATION123.

3.5.1 Learning paths

Learning paths can be understood as the **roadmap individuals follow during their learning journey.** This 'map' helps you to understand why learners decided to take part in a training, what decisions they had to make along the way, what difficulties they have encountered and how they dealt with them. It also allows us to see what other possibilities they had along the way or which directions they intend to take from now on.

All this information can be used to further **improve the trainings as well as their conditions you offer for their learning** journey. For example, information on the motivation or reasons to start a learning journey could help to cater and approach the learners better and it could help to reduce the drop-out rate and determine what actions you can take for the routes they intend to take in the close (subsequent to the training) and more distant future.

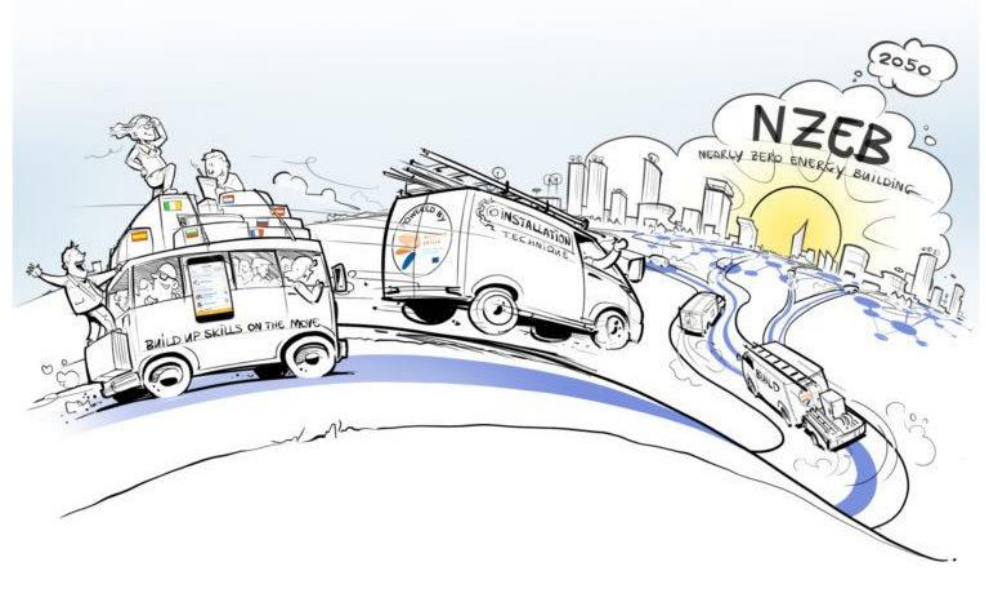


Fig 14: Understanding the route to success

The factors influencing adult learning are different from those affecting students. Personal matters, background, economic status, social factors, and even professional-related issues may be decisive for an adult to decide to begin, proceed with or quit a learning opportunity. So, if you want to increase the demand of the workforce for training and assure a successful transfer of learning, you need to identify what influences their motivation to engage in learning initiatives and how they continue afterwards.

The evaluation of learning paths includes the **understanding of how and why they decided to participate in a training, a description of their learning experience and what their expectations** are in relation to the knowledge and skills they acquired. This can be easily done using brief and simple survey instruments. In the EVALUATION123 – Question Pack you will find the survey Learning Paths ready for use (also in Appendix E).

Learning Paths in use – a case study

A BUSLeague partner organization offered upskilling training for energy advisors. Energy advisors are independent, highly specialized professionals that have only little opportunity to attend longer training sessions. Upskilling micro trainings were offered online.

To learn more about the focus group and understand what factors were influential in their learning processes an evaluation was implemented. The evaluation was designed according to the EVALUATION123 guide, and besides the reaction (Level 1) and Learning Outcomes (Level 2), it was decided to include also a few questions on Learning Paths.

The following questions were considered:

- *Through the training or the subsequent discussion, all my questions about the content were answered. (agreement scale: Totally disagree - totally agree)*
- *Which of your questions were not answered by the training?*
- *What will you do to clarify those questions?*
- *What will you do if you face a problem (regarding the topic of the training) at work?*

The overall results demonstrated that the training offered was of high quality and very relevant. The Reaction survey proved that the training was not only well received but also deemed very relevant, informing organization that the timing for offering the training was very precise. The Learning Goals assessment also demonstrated a significant increase in the perceived acquired knowledge by participants.

However, the Learning Paths survey revealed information that could not have been obtained by measuring satisfaction or skills only. The survey informed the organization that at least one third of trainees still had questions. A closer analysis in subsequent questions revealed that part of those questions related to topics not covered by the training. This informed what aspects should be reinforced (either in future communications or training) and what other questions and topics were recurrent among consultants. It was identified that, in order to solve their doubts or solve problems during their work practice, consultants would first search for information on the internet and secondly, discuss with colleagues. To a lesser extent, consultants would try to reach the organization or trainers for clarifications.

The questions that prompted participants to clarify their course-of-action when facing problems finally allowed the organization to understand what consultants need when applying their knowledge after the training in practice. This led to the consideration of follow-up meetings to address such problems and to strengthen the informal exchange between energy advisors for resolving related practical questions. In addition, acknowledging that the use of the internet is a very relevant factor, the organization is also able to consider other solutions for offering adequate and reliable support on the web.

3.5.2 Learning analytics

Learning analytics is **the collection, analysis, and reporting of data about learners and their environment for the purpose of understanding and improving learning outcomes**. By analysing and interpreting other sources of information than merely subjective and objective test results, namely (digital) traces of learners in their learning environments and their interaction with learning materials, learning analytics allows us to find opportunities for improving learning experiences like training and upskilling interventions. In the context of BUSLeague, for example**, understanding to what extent and why certain groups of workers are successful in completing a training program** is crucial for finding the best way of upskilling the sustainable energy skills of the workforce. To better illustrate this, let us think of a practical example.

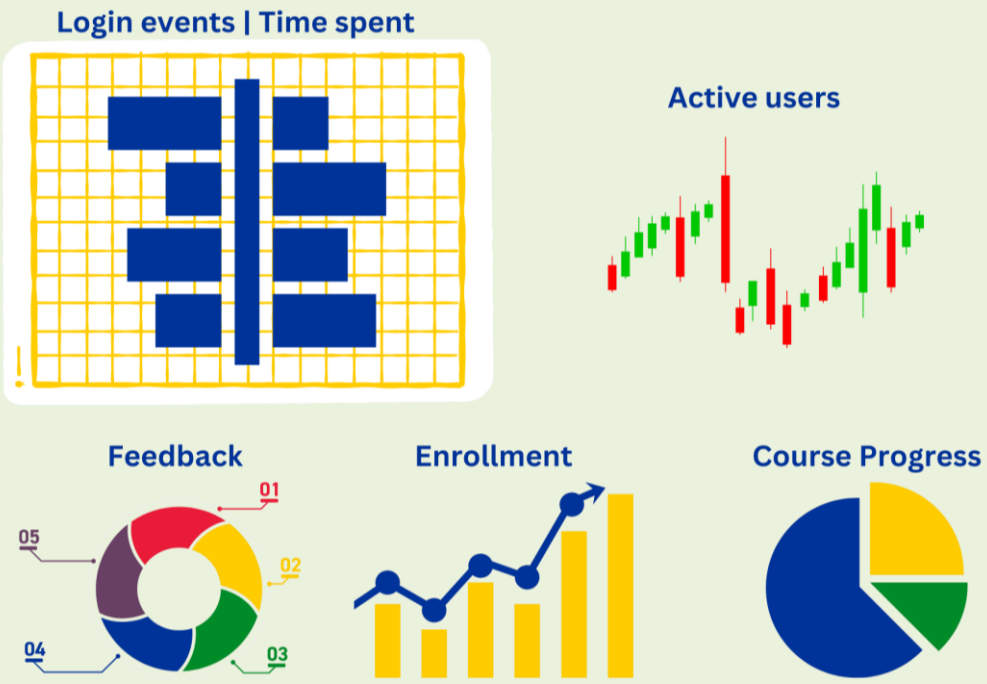


Fig 15: Understanding learning processes

Examples of questions that can be answered with learning analytics

First, imagine a hybrid training, consisting of face-to-face classes and online supporting material, such as videos, texts, and exercises. From the in-person part of the program, there are data that can be drawn upon, such as attendance, tardiness, participation during the classes, etc. From the online component, most platforms allow us to gather additional data, such as log-in instances, time spent on the platform, interaction with other users in forums, etc. Another source of information could derive from contextual factors, such as socio-demographic and professional data.

The combination of these data can produce a lot of interesting insights that you can use to improve our trainings and the learners' experience. Socio-demographic data, such as 'years of experience' and 'profession/ specifications', when combined with other data, may tell us whether the training is perceived as important or not for novices in comparison to experts of a certain professional practice, and allow us to react to in our instructional design. For example, attendance or usage data can inform us whether novice versus experienced electricians find a certain training relevant or challenging, and whether they have difficulties reconciling the training modes with their schedules.

These insights about the quality and efficacy of the learning material and conditions not only help instructors and developers of training programs but also provide arguments for learners to make a choice about dedicating their limited time to engage with learning materials or to participate in trainings and upskilling interventions.

*In a nutshell, you can say that learning analytics provide useful tools for **analysing**, **tracking** and **promoting** the **participation** of learners, their **progress** during the training and the **processes of knowledge creation** they engaged in.*

Most learning management systems already dispose of such capabilities and will provide certain insights based on analytics of digital data. However, you may also perform simple analogue data collection through attendance and performance sheets and enrich your understanding of the processes incurring alongside the training program, which will let you improve your trainings.

Implementing learning analytics with digital means

When implementing learning analytics for online trainings, learning analytics is enormously facilitated by technology. Most learning management systems and educational-related platforms already dispose of one or more analysis tools. These tools may already be integrated in the platforms, but they can also be available as extensions (plugins) or external tools like learning management platforms (e.g., Moodle). For example, some platforms provide reports on which contents were most used, visualized, or repeated (e.g., sections, simulations, videos, books, and links). This may imply that a specific content contains more relevant or complex information, which required more efforts from viewers. Other tools estimate how long a learner has dedicated to a specific part of the module or activity.

Make sure to check what the existing available tools are for your platform, and do not forget to perform a search on what other options might be available. In Appendix F you will find links to the exemplary tools for learning analytics in the Moodle platform.

Implementing learning analytics with non-digital means

In some instances, learning analytics can also be manually performed. By manually, we mean that the collection and input of data will not be automatized. Thus, the amount and quality of the data collected is heavily impacted. However, depending on your capability and resources, this might be more appropriate, and you would still benefit from all the information that this type of analysis provide. A manual attendance control sheet, for example, may provide you insights and allow you to cross data and make inferences on the underlying reasons for dropouts, or even reach out to them. A sample of a training attendance sheet, as well as a link for an online version, can be found in Appendix G.

It is also worth mentioning that this type of analysis and interpretation of data is only meaningful for voluntary and longer trainings, in which different measurements can be recorded over time.

Surveying dropouts

At times, you may have the opportunity to survey individuals who have not followed through a training. These are good opportunities to understand the reasons underlying this decision and identify possible constraints to the learning experience. A European study in adult education has identified six main reasons for dropping out:

1. **Organizational issues**: problems related to the organization of the educational experience, such as infrastructures, curriculum, bureaucracy, schedules, etc.
2. **Personal reasons**: problems associated to the family, finances, health, or other social factors that may interfere with the life of learners; also, other personal aspects such as performance anxiety, learning style and the levels of self-confidence and effort required influence a learner's decision to drop out a training or educational program.
3. **Skills**: problems related to specific skills required to participate in the training, such as languages, previous knowledge/experience, ICT, time management, etc.
4. **Goals**: lack of understanding, purpose or concrete notion of the benefits, opportunities and value associated with the training.

5. **Class dynamics**: problems related to the management of the class, involving the interaction among learners or between learners and instructors; common issues may be the poor differentiation between learners, failure to adapt contents and explanations to increase learners' comprehension, insufficient feedback or even in the creation and maintenance of interest by the instructors.
6. **Expectation**: problems related with the misalignment of the expectations of learners towards the training program; these may be represented in the methodology used, conflicting schedules, heavy workload during required for the training or the difficulty of the content of the program.

The best approach to survey dropouts and to assure responses is a quick, structured interview. In Appendix H we propose a 10-minute interview, with 5 guiding questions that will enable you to understand better the reasons underlying the decision of a participant to abandon a training.

Fig 16: Factors influencing Training Attrition

4 DEMONSTRATING EVALUATION DESIGNS AND INSIGHTS

The EVALUATION123 has been developed and evaluated as part of the one-and-a-half-year co-design trajectory with the complete BUSLeague consortium. We **demonstrate insights that have been gained by applying the EVALUATION123 reference guide and toolset**, with the help of one selected in-depth pilot and the beta-testing at the fifth consortium meeting.

This way, we can **illustrate actual results about trainee satisfaction, learning outcomes and future needs** (relation to learning paths) that were achieved through the evaluation framework and rubric, as well as **depict prominent evaluation plans that will be or have already been implemented by the partners of the BUSLeague consortium** for their upskilling interventions in the National Implementation Plans.

4.1 SUMMARIZED RESULTS OF SELECTED IN-DEPTH EVALUATION – AEA (AT)

The following section **describes the evaluation of the AEA pilot that focussed on the evaluation of a micro-training of energy advisors** on the topic of planning effective renovation roadmaps (see Figure 17). First, the **evaluation plan will be sketched** with the help of the EVALUATION123 Decision Tree Checklist. This is followed by selectively **reporting the results and the respective conclusion of the original report.** AEA was perceived as most insightful: usefulness of training, demographics, usefulness of assessing learning goals and quality of trainer. The learning goals assessment was capable of identifying LG neglected during the training and the satisfaction indicated that netEB was a good provider for the training.

Fig 17: AEA Report on results of EVALUATION123 pilot at the fifth consortium meeting.

4.1.1 Evaluation plan – Decision Tree Checklist

The training "The construction site process – What influence does it have on the creation of the renovation roadmap?" was commissioned by the Austrian Energy Agency to Netzwerk Energieberatung Steiermark and offered to energy advisors. These advisors are self-employed and take part in these network trainings voluntarily. Their interest in networking drives their

motivation in participation, aside from gaining new knowledge for their job. The trainer is one of the rare individuals able to deliver a training to the focused topical experts, the advisors, and therefore represents an irreplaceable resource for AEA as well as the NEB. She or he is not always paid for sharing the expertise as part of such events, limiting the options of asking for contribution to an evaluation and providing additional services such as slides of the talk.

The intervention took place in December 2021, as part of a network meeting held by AEA. The **evaluation plan is traced and depicted in the form of the EVALUATION123 Decision Tree Checklist (see Figure 18).** The goal of the evaluation was to get the most out of a Level 1 & 2 evaluation of Kirkpatrick, i.e. understanding the reactions and learnings of the participating advisors, without throwing them off (increased dropout of training). Approximately 30 people took part in the training and were prompted to reply to a survey. A total of 11 respondents completed the survey. Due to the voluntary participation in the network meetings, there is no chance for a follow up evaluation.

UNDERSTANDING THE TYPE OF TRAINING

Institution or Organization	**Contact person**
AEA	**Georg Trnka**

Training
The construction site process – What influence does it have on the creation of renovation roadmap? Date: 13. December 2021, 15:00 to 17:00
Focus of training Practical training: The construction site process – What influence does it have on the preparation of renovation schedules? • Processes and responsibilities on the construction site • Correct communication between the trades • What is a rehabilitation roadmap and how can the renovation work in steps? • Practical examples DI DI(FH) Ing. Harald Reiter (Reiter GmbH)

Target audience:	***Number of Participants:***
Energy consultants	30

Provider of training	
☐ Own institution or organization ☑ Commissioned training provider: *Netzwerk Energieberatung Steiermark*	☐ Partner institution or organization ☐ Other:

Type of Training		☑ Voluntary ☐ Mandatory
☐ In-class ☐ On-site ☑ Online	☐ Hybrid ☐ Other:	**Duration and length of training** 2-hour online workshop

Elements contained in the training program (multiple items may apply)

☐ face-to-face instruction ☐ reading material ☐ hands-on practice	☑ Q&A session ☐ demonstration ☐ other:	**Digital** ☐ video lessons ☐ e-learning ☐ discussion forum	☑ online tutorial ☐ Other:

Additional information/ comments

- Training was offered as part of a regular scheduled network meeting for energy consultants
- Content and training provided by Netzwerk Energieberatung Steiermark

Structure: post-test only, no knowledge assessment

Post-test: post-test was designed according to Eval123

- Platform: Qualtrics
- **Level 1**: socio-demographics, satisfaction with training, quality of training, opportunity to interact and relevance of training
- **Level 2**: Self-reported score on LGpre and LGpost
- Indicators for **Level 3**: motivation to transfer knowledge, resolution of training related doubts, resolution of problems, interest in exchanging ideas with peers,

Fig 18: Short form of evaluation plan of AEA in the first pilot traced with the Decision Tree Checklist.

4.1.2 Interpretation of results

This section summarises the data collected in evaluation of the micro-training. The training was commissioned to the Netzwerk Energieberatung Steiermark, a training provider, by the Austrian Energy Agency and offered to energy advisors. The intervention took place in December 2021, as part of a network meeting held by AEA.

The evaluation was a pilot within task 5.3 of the BUSLeague project to apply the rubric for measuring educational effectiveness and explore what is missing for guiding the evaluation of (micro-)training interventions in the realm of the energy transition. In the following the key statistics and their interpretation will be described. Details on the evaluation can be found in the appendix 7.4.

Demographics

- A total of 11 participants responded to the survey. The rate of response was 100% for quantitative questions.
- 54% of respondents were aged 35-54 years old.
- 82% had a university diploma.
- 54% of respondents had at least 5 years of experience as Energy Advisors.
- 64% of respondents had previous experience in the energy or construction sector, with an average of 9 years of practice.

Learning Goals

Self-reported score on perception of mastery of learning goals BEFORE the training:

Minimum	Average	Maximum
42.40	**71.75**	99

Self-reported score on perception of mastery of learning goals AFTER the training:

Minimum	Average	Maximum
48.50	**78.77**	100

Net average gain on perception of achievement of learning goals:

Minimum	Average	Maximum
-6.30	**7.32**	18.94

Respondents scored on average 7.32 percentual points higher in reported learning goals achievement.

The learning goals with highest achievement scores (>10%) were:

- Explain the most important factors for the successful implementation of a renovation roadmap.
- Ability to independently sketch a renovation roadmap.

Satisfaction with Training

82% of respondents were very satisfied with the training in terms of motivation to participate and learning opportunity. Respondents did not feel particularly encouraged to pursue additional reading or do further research on the topics presented.

Quality of Training

On average, respondents were very satisfied with the quality of the instructor and the training provided (4.74 out of 5), with very little variance in responses.

Quality of Interaction

The average rating for the quality of the interaction during training was 3.79 on a 1-5 scale. Respondents reported lack of opportunity to make acquaintance and interact with their peers.

Relevance

All respondents rated the contents of the training and the associated learning objectives as very relevant for their work as energy consultants.

Questions about the training

27.3% of respondents had questions which were not answered by the training. 75% of respondents reported that they will discuss their doubts with colleagues or search for the response. None of the respondents considered contacting NES or the trainer.

Motivation

91% of respondents fully agreed and 9% agreed that they feel able to apply what they have learned in their practice as energy consultants.

Troubleshooting

When facing a work-related problem, all respondents reported to search for an answer, either on the internet or elsewhere. 82% reported to also discuss with their colleagues when solving a problem. 55% claimed to gather experience form practice and consult with NES.

Follow-up

45.5% of respondents reported their wish to exchange ideas and experiences with the topics discussed with their network colleagues. 55.5% remained neutral.

4.2 EVALUATION CONTEXTS AND PLANS OF FUTURE EVALUATIONS – BL CONSORTIUM

As part of the last workshop at the fifth consortium meeting, the nine BUSLeague partners could leverage the complete EVALUATION123 to plan their upcoming upskilling activities: IGBC , TUS, IVE, BAUHAUS, AVE, ISSO, AEA, IRI-UL & BCC. This section **describes the partners' anticipated evaluation context first of the BUSLeague partners first, and then explains the resulting most prominent evaluation plans**. These are traced and depicted with the help of the EVALUATION123 Decision Tree Checklist again.

4.2.1 Anticipated evaluation context

In total, 9 evaluations were planned and respective evaluation contexts are listed in Table 5. Only, IRI-UL was not considered in the analysis as they did not represent an application partner but the ethnographic perspective on evaluation as part of WP5.

From the 8 planned evaluations, we learned that ~40% are developed by their own institution and the rest by partner organisations or are commissioned to third parties. 75% of trainings had online components, or took place completely online. 50% of the trainings claimed to use some sort of e-learning platform. ~50% of the trainings were offered voluntarily to professionals. They endure up to 30 minutes in ~40% of the trainings, while two trainings were expanding over a longer timeframe with several modules. Further, hands-on practice was reported in 40% of the trainings. Video lessons, reading materials and forums were mentioned as learning materials.

Partner	Provider of training	Type of Training	Atten-dance	Duration	Elements
AEA	Commissioned training provider	Online	Voluntary	10-20 min	video lessons, e-learning platform
BAUHAUS	Partner institution or organization	Online	Voluntary	30 min	video lessons, e-learning platform, reading material
BCC	Own institution or organization	Online	Mandatory	6-7 months	e-learning platform
ISSO	Partner institution or organization,	Hybrid	Mandatory		e-learning platform, hands-on practice, f2f
IVE	Partner institution or organization	Online		30 min	video lessons, e-learning platform, reading material
TUS	Commissioned training provider	Hybrid	Voluntary	—	online tutorial, e-learning platform, hands-on practice, f2f, reading material, forum
AVE	Commissioned training provider	In-class or on-site	Mandatory	12h in 4 modules	hands-on practice
IGBC	Own institution or organization	In-class or on-site, Online	Voluntary	—	f2f

Table 5: Overview of evaluation contexts of the 9 plans of upcoming upskilling activities.

4.2.2 *Prominent evaluation designs*

In terms of the designed evaluation plans, we learned that **all partners are able to survey trainees and perform a Level 1 – reaction to learning survey**. Only one organisation mentioned to perform both quality assessment via the design principles and the reaction survey. The finding that quality assessment is not conducted by all partners might be reasoned in the Decision Tree Graphic. Apparently, it did not make clear that following both paths is possible. This is corrected in later versions of the EVALUATION123. **~60% of the trainings by the BUSLeague consortium did not aim for a certification, but did continue with planning evaluations on level 2 and 3**, i.e. considering learning from training and changes in behaviour due to training. Only two trainings claimed to be able to perform both pre and post-tests. Finally, **three trainings indicated interest in performing evaluation level**, i.e. incorporate an evaluation of the transfer of learning.

There are **two prominent evaluation designs, i.e. patterns of evaluation design decisions that are followed up by parts of the consortium**. The partners are roughly distributed equally on both patterns, with one partner being indecisive and all partners having their specificities. We consider **pattern 1 to be the more realistic pattern as evaluations of upskilling activities are often required to be compact**, and this pattern can be conducted directly after an evaluation in one run. This is consistent with the learnings from the pilots as well as consortium workshops.

Both patterns can be differentiated based on the inherent evaluation goals, level of evaluation and requirements for evaluation:

1) **Pattern → see Figure 19:**
    a. About 4 representatives
    b. Level of evaluation: includes
        i. Level 1 - Reaction to training
        ii. Level 2 - Learning from training
    c. Most of representatives include only post-test
    d. (Interest in considering learning paths and use learning analytics)

2) **Pattern → see Figure 20:**
    a. About 3 representatives
    b. Level of evaluation: includes
        i. Level 1 - Reaction to training
        ii. Level 2 - Learning from training
        iii. Level 3 - Changes in behaviour due to training
    c. Most of representatives include only post-test
    d. (No interest in considering learning paths and use learning analytics)

Please, mind that these evaluation plans have been conducted with older versions of the EVALUATION123. Aside from the issue with underlining the importance of quality assessment in each evaluation, the **Decision Tree did also neglect learning analytics and learning paths** and not clearly trigger respective reflections. The **interest in learning paths and analytics therefore has to be interpreted carefully.** This consortium workshop was used to receive final feedback and improve on the toolset. See the previous chapter or the EVALUATION123 for the final versions.

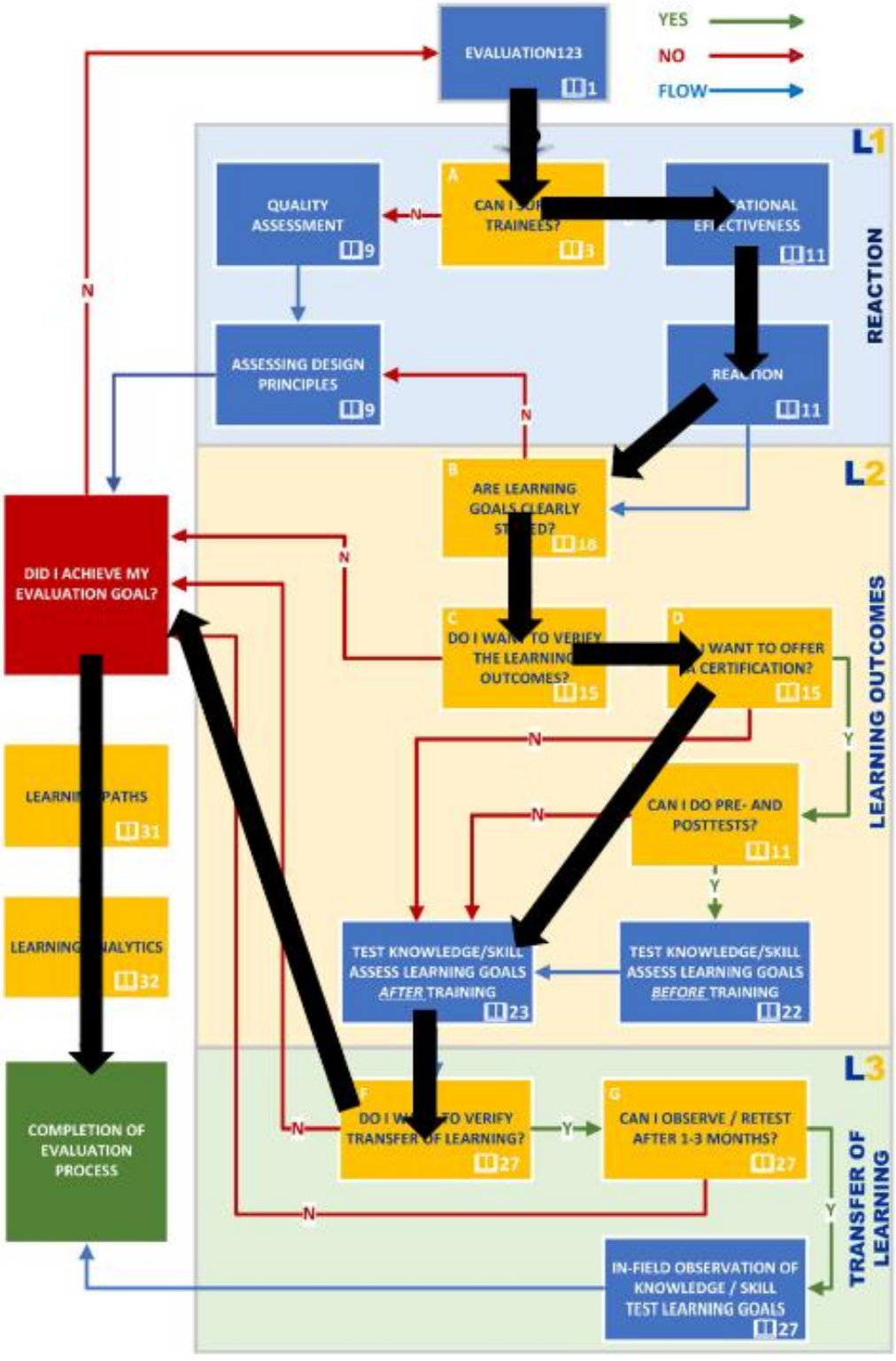


Fig 19: Prominent evaluation plan in the BUSLeague consortium - first pattern ends at level 2.

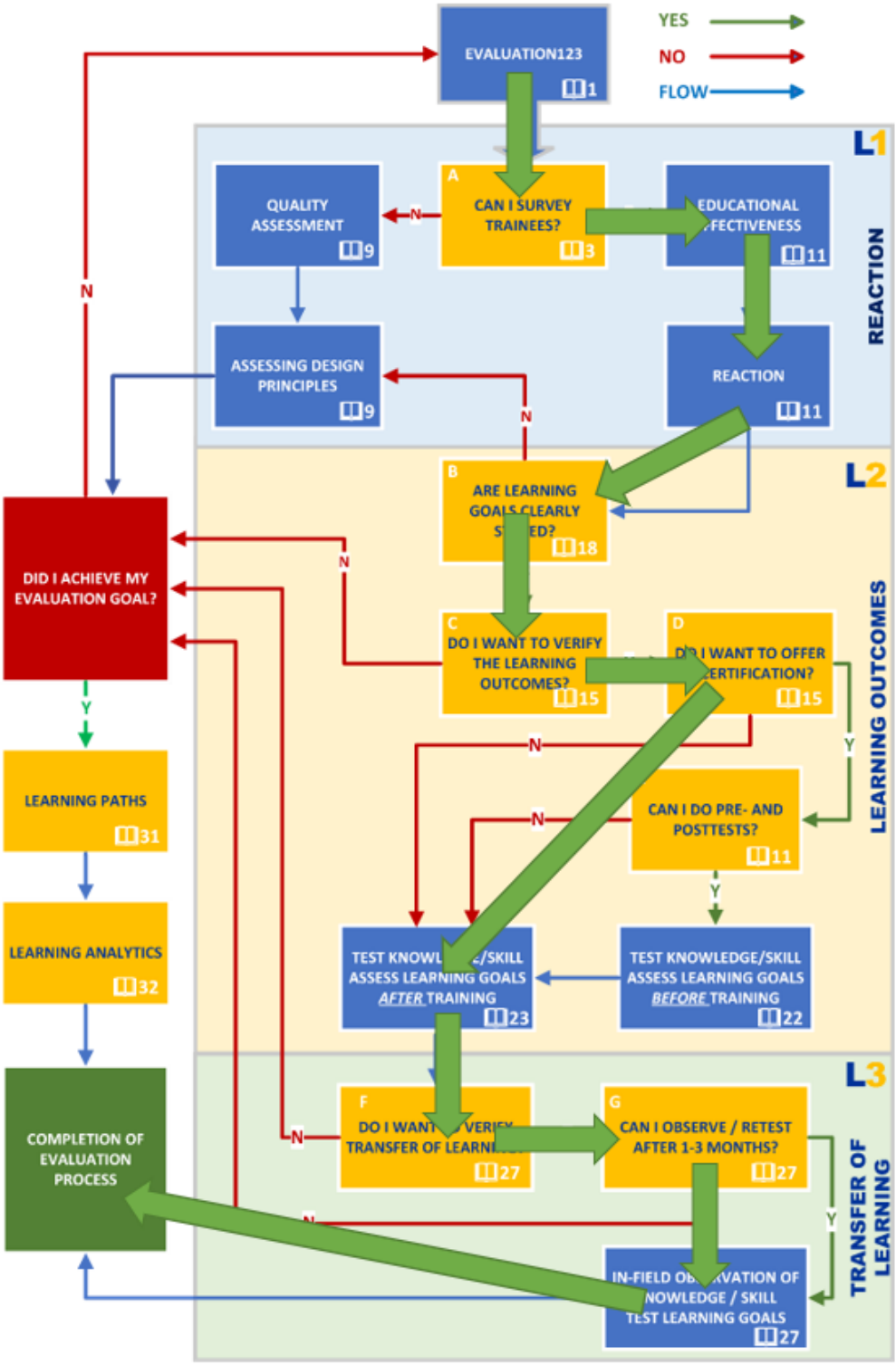


Fig 20: Prominent evaluation plan in the BUSLeague consortium - second pattern including level 3.

5 CONCLUSION

The BUSLeague consortium has embarked on the mission to provide upskilling interventions for the energy transition and to address the demand for a skilled workforce. **To understand the effectiveness of their evaluations, the partners developed evaluations with the help of the EVALUATION123** that consider their evaluation goals, ranging from improving the upskilling intervention to providing certifications of the acquired knowledge and skills. **Reflection on the trade-off between evaluation benefits and efforts with the decision tree** thereby helped them to determine the best evaluation design for their evaluation context and respective limitations.

For designing suitable interventions and evaluating their effectiveness, the **EVALUATION123 motivated the partners to systematically reflect on the learning outcomes** that they wanted to address and the outcomes that they wanted to achieve. This included transferring general learning outcomes from predefined ULOs into **specific measurable learning goals with the help of the learning goal tool.** In addition to their positive impact on self-regulated professional learning via self-assessment, the explication of learning goals helped to disclose specific overlapping upskilling objectives in the consortium. This paved the way to **meaningfully discuss and share evaluation plans (checklist) and approaches (questionnaire packs and items)**, together with underlying training materials. In addition, the systematic definition of learning goals holds the potential to compare effectiveness of different upskilling interventions for the multifaceted construction professionals in the future.

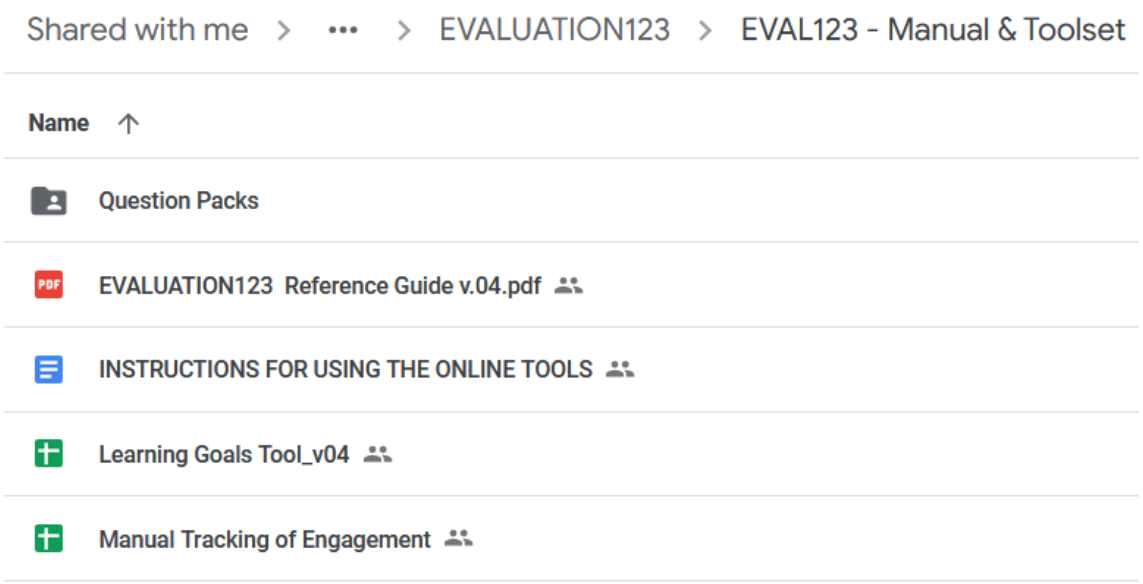


Fig 21: Google Drive Public Access to the EVALUATION123 reference guide and toolset.

The EVALUATION123 reference guide and tools are **publicly available via Google Drive for cost-free adoption for all interested stakeholder in the construction and building domain**: https://drive.google.com/drive/folders/1sKn0ZBdj__neC8BdCpqW_d9NcabfCrM3). The Google Drive folder also provides a **platform for sharing evaluation plans after project runtime, and offers to reflect and discuss** the EVALUATION123 for future adaptation: https://drive.google.com/drive/folders/1NykZAJ25G_RzWtxvBrCYC8bVCjefquxX . Besides, the open access pre-publication (https://osf.io/preprints/socarxiv/yb8w5/) enables adoption beyond the Google Suite within inhouse software environments.

The EVALUATION123 equipped the partners of the BUSLeague consortium with an **easy to read introductory handbook, the reference guide, explaining and referring to the toolset to simplify its adoption and scaffold the evaluation of effectiveness**. The appreciation of this step-by-step guidance is reflected in its uptake by related EU-projects already during project runtime (e.g. BUS Go Circular) and a comment of an BUSLeague consortium partner: "The Evaluation123 is a great opportunity to improve our work in providing trainings for construction and building professionals."

In line with review studies on the status of effectiveness research (see Chapter 1), the professionals' reaction to and learning from training (Level 1 & 2) have been in the focus as part of the co-design trajectory of the EVALUATION123 - except one pilot leveraging subjective indicators for transfer of learning. However, in current and upcoming evaluations of upskilling interventions, about half of the BUS League consortium aimed at evaluating effectiveness also objectively on level 3, i.e. they are interested in assessing changes in behaviour due to training. This is reflected in the prominent evaluation plans that were identified at the fifth consortium meeting. In these plans and parallel discussion, half of the consortium showed interest in learning paths and analytics in the future evaluations, as well. This number might even be higher as the Decision Tree missed to explicitly trigger respective reflections in earlier versions.

The ambitions of the BUS league consortium paint a promising picture for the future upskilling interventions as they reflect **interest in the workplace learning approach to learning and evaluation**. Partners do not only look at individual knowledge acquisition, but also successful socialisation and joint knowledge creation. Understanding **learning paths provides insights into open needs of trainees such as providing chances for community exchang**e on remaining questions or issues in practice after an upskilling intervention. On the other hand, it was discussed to **extend the BUS-APP with features to record and share learning moments subsequent to an upskilling intervention**. This could not only serve as evidence for transfer of learning, but also promote the exchange and knowledge creation between professionals in solving actual issues of daily practice.

6 ACKNOWLEDGMENT

T5.4 conducted a design research trajectory in collaboration with the complete BUS League consortium. We sincerely thank all partners for their active and constructive engagement in the pilots and consortium workshops to co-design the EVALUATION123. Furthermore, we thank Loes de Jong for her contribution in the early phases of the BUS League project and setting the scene for the co-design trajectory, as well as Sandra Schele for polishing the appearance of the deliverable in the final stages. And we especially thank Leandro Hirose for his wholehearted engagement in bringing forward the EVALUATION123 and re-designing reference guide and toolset with perseverance so that it finally meets the needs of the BUS League consortium.

7 APPENDIX

7.1 DETAILED RESULTS OF LITERATURE STUDY

Predictors of training transfer

Research on transfer of training distinguishes between individual elements (e.g. attitudes toward training), job and career elements (e.g. decision-making power), situational elements (e.g. organisation support), training elements (e.g. design and feedback), training outcomes (e.g. skill acquisition) and motivation to transfer, that together predict training transfer. An overview of all categories can be found in Table 6.

Table 6: Influencing elements of transfer of training

Influencing element (main references)	Examples
Individual elements (Cheng & Hampson, 2008)	Locus of control Conscientiousness Anxiety Goal orientation Attitudes toward training (Burke & Hutchins, 2008; Laker & Powel, 2011) Motivation to learn (Gegenfurtner et al., 2009; Burke & Hutchins, 2008) Personality traits (Gegenfurtner et al., 2009; Burke & Hutchins, 2008) Work commitment (similar to career commitment; Gegenfurtner et al., 2009) Ability (Burke & Hutchins, 2008) Perceptions (Burke & Hutchins, 2008) Expectations (Burke & Hutchins, 2008) Prior learning experiences (Laker & Powell, 2011)
Job/career elements (Cheng & Hampson, 2008)	Job involvement Organizational commitment, organizational climate (Festner & Gruber, 2008), organizational culture (Gegenfurtner et al., 2009) Career commitment Work load (Russ-Eft, 2002; Festner & Gruber, 2008) Decision making power (Festner & Gruber, 2008)
Situational elements (Cheng & Hampson, 2008)	Opportunity to transfer Transfer climate (supervisor support, supervisor sanctions, peer support; Russ-Eft, 2002; Laker & Powell, 2011; organization support; Burke & Hutchins, 2008; Laker & Powell, 2011) Intervention strategies (similar to post training)
Training elements (Russ-Eft, 2002; Saks & Belcourt, 2006)	*Pre-training* (also Gegenfurtner et al., 2009) Persuasive message linking mastery and job survival Realistic training previews Voluntary versus involuntary Trainee preparation Trainee input and involvement Supervisor involvement Training attendance policy Needs assessment information (Burke & Hutchins, 2008; Laker & Powel, 2011) Training design - Advance organizers

	- Guided discovery - Error-based learning versus digital - Metacognitive instruction - Learner control - Mastery orientation versus performance orientation - Behavioural practice versus symbolic - Spaced practice - Variable examples/practice - Random practice - Overlearning - Coaching/feedback/scaffolding - Identical elements (also Laker & Powell, 2011) - Stimulus variability - General principles - Feedback - Positive reinforcement - Goal setting (also Burke & Hutchins, 2008; Laker & Powell, 2011) - Relapse prevention - Focus on soft or hard skills (Laker & Powell, 2011) - Learning activities (Burke & Hutchins, 2008) - Trainer support (Burke & Hutchins, 2008) - Immediacy and feedback (Laker & Powell, 2011) *Post training (also Cheng & Hampson, 2008)* Goal setting; proximal Relapse prevention training Self-management Training in self-talk Visualisation Post training follow-up Supervisor support (similar to transfer climate) Organization support (similar to transfer climate) Accountability Evaluation and feedback
Motivation to transfer (Cheng & Hampson, 2008	Extent to which trainees are motivated to transfer learnings is expected to influence transfer behaviour. Transfer motivation in turn is influenced by training outcomes, and individual factors (individual characteristics, job/career variables & situational variables).
Training outcomes (Cheng & Hampson, 2008)	Post training self-efficacy Reaction to training Declarative knowledge Skill acquisition (Gegenfurtner et al., 2009) Self-efficacy (Gegenfurtner et al., 2009; Laker & Powell, 2011) Expectations (Gegenfurtner et al., 2009) Training reactions (Gegenfurtner et al., 2009) Level of expertise (Laker & Powell, 2011)

7.1.1 Measuring effectiveness of upskilling

Suitable proposals for measuring the educational effectiveness of upskilling interventions are extracted from the literature study. Table 7 lists these proposals according to what and how they are measuring the effectiveness, first from the perspective of conventional approach to learning and then for the workplace learning perspective.

Table 7: Proposals of measuring effectiveness (what and how)

Author(s)	Measuring what	Measuring how
Conventional approach to learning		
Cheng & Hampson (2008) – Literature review	Transfer of training - Individual characteristics - Job/career variables - Situational variables - Motivation to transfer - Training outcomes: "Yet, training needs to be demonstrably effective. Kirkpatrick's (1967) taxonomy evaluated training on four 'levels': according to how trainees felt about it (i.e. reactions), whether they learned anything (i.e. learning), whether the learning was transferred to the job (i.e. behaviour) and whether it achieved its objectives, in terms of fixing the perceived performance deficit that was the training's rationale in the first place (i.e. results)."	
Cant & Levett-Jones (2021) – Literature review	"In order to describe the level of educational impact we classified the results of each review according to Kirkpatrick's Four Level Model of training evaluation (Kirkpatrick, 1975, 2006). The four levels of evaluation interpreted for evaluation in healthcare are: • Level 1. Reaction (indicated by program evaluations such as participant satisfaction) • Level 2. Learning (knowledge tests, performance scores) • Level 3. Behaviour (transfer to practice - change in clinical practice) • Level 4. Results (impact on service delivery or on patient care)."	
Russ-Eft (2002) – Literature review	Transfer of training: - Situational elements - Pretraining elements - Training design elements - Post training elements	
Saks & Belcourt (2006)	Transfer of training: - Pretraining activities - Activities during training - Post-training activities	Questionnaire
Gao et al. (2019)	Training in construction sector - Knowledge - Acquisition - Behaviour alteration - Injury rate reduction	- Self-reports - Supervisor reports

Oliver et al. (2004)	Bloom's taxonomy for analysing the cognitive depth of performing a given task.	Assessment
Cunningham (2007)	Satisfaction with activity	Questionnaire
Athiyaman (1997)	Satisfaction with course	Questionnaire
Chauhan et al. (2016)	Transfer of training	Questionnaire
Workplace learning		
Tynjälä (2013) – Literature review	3-P model of workplace learning: Presage: - Learner factors (and interpretation) - Learning context (and interpretation) Process: - Activities Product: - Learning outcomes	
Park et al. (2021) – Literature review	- Antecedents - Mediator/moderator - Outcomes	Quantitative studies
	- Participation or engagement in workplace learning activities - Workplace learning by	- Respondents' self-perceptions (indicate participation in activities on a Likert-scale; yes or no; number of days; number of competencies; allocation of 100 points
	- Learning outcomes	- Performance score
	- Professional learning	- Evaluation of instructors (objective) - Self-rated evaluation (subjective)
Park & Lee (2018)	- Presage - Process - Product	Questionnaires
Ruiz-Calleja et al. (2017) – Literature review	Learning evaluation: - Learning process - Tracking development of competencies for reflection, group assessment or self-assessment Technical evaluation Acceptance evaluation	Who: - Learners/workers - Learners/apprentices - Trainers - Managers Data: - Synthetic logs - Authentic logs - User feedback - Observation - Learner-produced artefacts - Demographic data - Recordings - Interviews - Questionnaires - Physical data

		Through: - Application - Same app. - Infrastructure - Same inf. - VLE - MOOC - Other platform With: - Social network - Discourse analysis - Content analysis - Context analysis
Schaaf et al. (2014)	Workplace learning with e-portfolio Competencies: - Knowledge - Skills - Attitudes	- Performance - Observable behaviour
E-learning		
Chao et al. (2009)	E-learning criteria	Questionnaire
Tzeng et al. (2007)	E-learning criteria	Questionnaire
Wang (2003)	E-learning criteria	Questionnaire
Wei et al. (2021)	Learning outcomes of MOOCS Cognitive outcomes: - Knowledge - Intellectual skills	- Course requirements - Discussion forum - Quiz - Exam - Assignment - Survey
	Behavioural outcomes: - Engagement - Course completion	- Survey - User data - Observation
	Affective outcomes: - Perceptions of MOOCS experience - Perceptions of MOOCS benefits - Course satisfaction	- Survey - Interview - Observation
Other		
Kelly et al. (2009)	Student attitudes to learning skills through video	- Questionnaire

7.2 DESCRIPTION OF THE CO-DESIGN TRAJECTORY

Each of these co-design activities will be summarised in the following subsections delineating the main lessons learned about evaluation in the construction and building domain, and the design requirement for the EVALUATION123, consisting out of a handbook and auxiliary tools.

Pilot 1 – Austria (AT)

The first pilot was conducted with the "Austrian Energy Agency" (AEA) and focussed on the evaluation of a micro-training of energy advisors on the topic of planning effective renovation roadmaps. AEA commissioned the Styrian network for energy advisors to find a speaker for the 2-hour online training as part of their regular voluntary network meetings. 30 energy advisors participated in the event and included training.

The evaluation should inform the future provision of training in cooperation with the provider, the suitability of the trainer and the relevancy of the selected training contents. The **evaluation design of the micro-training for energy advisors was guided by the evaluation framework and rubric and provided insights into their value and shortcomings**. AEA was eager to assess all three levels of evaluation, but practical constraints did not allow implementing a pre-test for level 2, a knowledge test or follow up measurements for level 3. As the trainer was not available for assessment, the trainees could only assess their achievement of the learning goals themselves (self-evaluation - subjective measurement). These goals were not existent and inferred from the learning contents.

The evaluation was conducted in the form of a post test that was distributed with the fee-based online survey tool "Qualtrics" after the training. It comprised of the elements:

- Socio-demographics
- Level 1 - subjective measurement (subscales): satisfaction with training, quality of training, opportunity to interact and relevance of training
- Level 2 – subjective measurement: Self-reported achievement of learning goals before and after the training
- Indicators for Level 3 - subjective measurement: motivation to transfer knowledge, resolution of questions and practical problems, interest in exchanging ideas with peers

As insights in level 3 were desired, but could not be evaluated in a follow up measurement, the motivation to transfer the gained knowledge and skills was included as an indicator. In terms of learning paths, furthermore, ways of resolving training related doubts and later problems in practice were explored, together with the interest in exchanging with peers after the event.

AEA considered the evaluation via the subjective measurements of reaction and learning as appropriate in terms of the initial goals, and underlined the value of the learning goal achievement indicated through trainee's self-assessment. Results provided insight into the usefulness of the training from the perspective energy advisor, perceived gaps of the trainer in addressing the promised learning contents and remaining learning needs subsequent to the training. The evaluation process was perceived as demanding and time-consuming in terms of the evaluation design, implementation and analysis. Especially in terms of data analysis, resources were limited. Out of this reason, AEA preferred automatic reports on descriptive statistics via the survey tool Qualtrics but in a costless version. Furthermore, AEA considered the definition of systematic learning goals as complex, especially with respect to defining specific and measurable goals. The rubric alone is not sufficient to guide evaluations given the limited time and resources. The partner saw the need in supporting the definition of learning goals, the selection of instruments from the rubric and structuring an evaluation in general.

7.2.1 Pilot 2 – Spain (ES)

We conducted the second pilot with a Spanish DIY-store of the Bauhaus retail chain. The evaluation focussed on a training of the sales staff. It should increase their literacy in energy efficiency and their ability to assist customers in making appropriate decisions when purchasing renovation or construction materials. The training consisted of 10 videos and additional materials that were provided in a module of the learning management system MOODLE. The

participation of the 1250 employees was mandatory, and had to be completed successfully. Unstandardized instruments were already in place for surveying satisfaction with the training.

The evaluation should clarify what knowledge the trainees learned and how this knowledge translates into the working practice. Again, the **evaluation framework and rubric guided the design of the evaluation to further elaborate on their value and shortcomings**. Even though there was interest in the transfer of learnings into practice, Bauhaus wanted to receive recommendation for level 3 and evaluate the levels 1 & 2 in the first instance only. For level 1 & 2, a pre- and post-test could be conducted as well as a knowledge test for level two. Learning goals were not existent and inferred from the videos on literacy in energy efficiency.

The evaluation was conducted with an online questionnaire via the MOODLE platform before and after the employees had conducted the training module. It comprised of the elements:

- Level 1: satisfaction with a custom MOODLE questionnaire by Bauhaus
- Level 2:
    - Objective measurement – Designed knowledge test in MOODLE questionnaire
    - Subjective measurement - Self-reported achievement of learning goals in Gform

For the analysis, Bauhaus could make available data of the subjective achievement of learning goal and discuss these results in relation to their own analysis of other data of level 1 and 2.

Bauhaus perceived the evaluation goal of assessing the knowledge acquired by their employees addressed by the objective and subjective measurement of learning outcomes. Thereby, results of the subjective assessments of the learning goals were reported to confirm the results of the objective knowledge tests. This increased the appreciation of measuring learning goals. Except for the implementation of the evaluation with MOODLE, the evaluation process was again perceived as demanding and time-consuming in terms of the evaluation design and analysis. Bauhaus considered the definition of systematic learning goals as well as the elaboration of the knowledge test as difficult. The rubric was perceived as too complex as it did not explain the used terminology in lay-man's terms and the relevancy of entailed scales. The partner saw the need to simplify and explain the rubric and support the definition of learning goals.

7.2.2 Concluding on EVALUATION123 after Insights from first 2 Pilots

Bauhaus and AEA were able to fulfil the stated evaluation goals based on the structure and instruments underlying the evaluation framework and its rubric. Our assistance in their application was central, however, and revealed severe problems for the self-directed application. Both partners did not perceive the evaluation framework and rubric as easy to use. This was attributed to its complex terminology, unclarity of the evaluation levels, missing explanation of instruments and missing support in defining systematic learning goals. The latter learning goals were not existent at the start of the pilots. They were perceived helpful in efficiently evaluating trainings via subjective measurements, and informing training development.

We concluded that the **EVALUATION123 must provide a step-by-step guidance for understanding the evaluation framework and rubric: (1) terminology, (2) structure of levels, (3) related instruments or scales and (4) their application**. The framework and its toolset need to be simple, usable and valuable in terms of predefined evaluation goals so that it can be implemented autonomously even by inexperienced evaluators. Otherwise scarcity of time, human and material resources restricts its adoption in practice. And the **definition of learning goals needs to be assisted**, in order to produce clear, viable and measurable training objectives. Their self-assessment could be leveraged in evaluation contexts that do not allow for any objective measurement on evaluation level 2 (learning from training) and level 3 (changes in behaviour due to training).

These partial design requirements resulted in the definition of **design propositions that resulted in three design artefacts (see Section 3.3 for final EVALUATION123)**:

5) **Reference Guide (digital handbook/explanatory manual)**: Explain concepts needed to design an evaluation with the evaluation framework - quickly consultable and referable; depict purpose and functioning of an evaluation in a simplified visual way; differentiate between subjective and objective measures on all three levels of evaluation; guide users step-by-step, on what needs to be done when with which tools.
6) **Learning Goals Tool (Google Spreadsheet)**: Explain the definition of specific and measurable learning goals following Bloom's systematic on evaluation level 2 & 3; support the systematic formulation of learning goal elements by providing partial structure, formulations and examples.
7) **Question Packs (Google Forms)**: Separate the instruments of the rubric into meaningful sets that can be easily replicated and combined into new surveys; allow self-assessment of learning goals on level 2 & 3; propose free and easy to use survey software that automatically produce descriptive statistics.

7.2.3 Consortium Workshop 1 – NL, FR, AT, IE & ES

The first co-design workshop on consortium level involved 9 participants from 8 partners, namely ISSO (NL), Practee Formations (FR), AVE (FR), AEA (AT), IGBC (IE), TUS (IE), IVE (ES) & Bauhaus (ES). We aimed at **introducing the new design artefacts, the learning goals tool and the question packs, and receiving feedback on them**. The question was how both tools can support the self-directed application of the evaluation framework and rubric and what is feasible and useful or needs to be improved. Accordingly, the workshop was split in two activities based on a fictive evaluation case (see Figure 21) to ease collaboration across partners in four mixed groups: 1) Definition of two learning goals with the help of learning goals tool & 2) Creation of questionnaire (incl. learning goals assessment) with the question packs. In addition, the first version of the **reference guide was presented and discussed at the end of the workshop**, as the explanatory manual that will link the tools to the evaluation framework and provide instructions for their application.

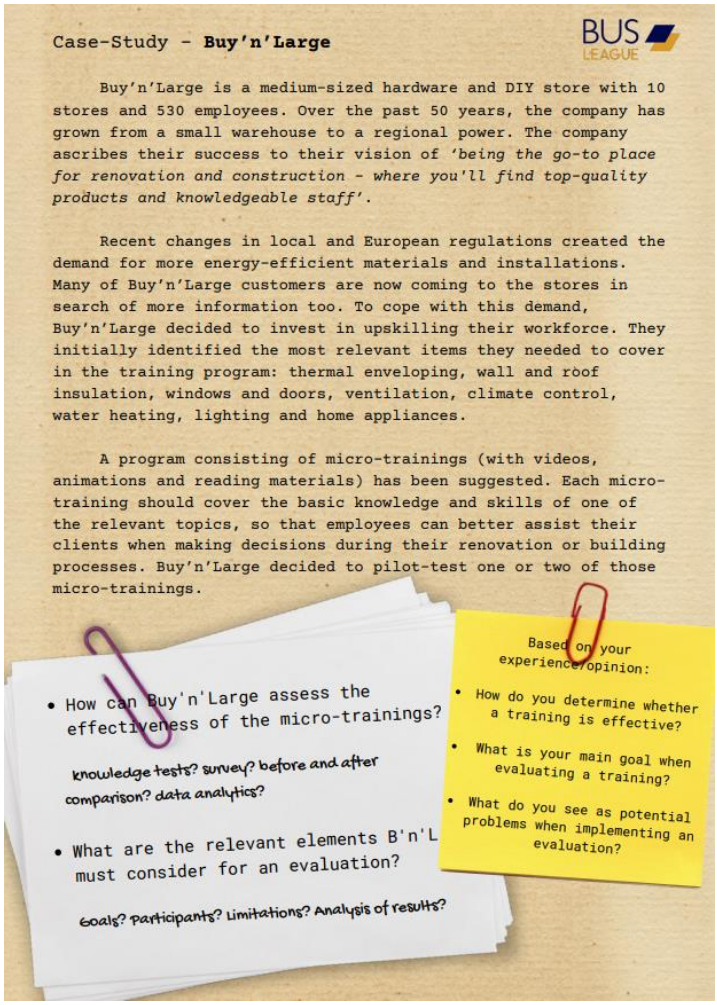


Fig 21: Description of the fictive case for the co-design activity including preparatory questions.

With respect to the first activity, all four groups saw the learning goal tool as useful in defining specific learning goals autonomously. The tool helped them to narrow down a general unit of learning outcomes (ULO) into a measurable action. The partners criticised the lack of overall explanation, the used terminology being too complex, not all verbs being suitable and too many options to choose from. In addition, they desired a translation feature to ease the formulation of learning goals in their mother tongue. One participant also recognized the importance of the learning goals beyond the self-assessment for the instruction of the learner and specifying training contents: "I also think [these learning goals] make a lot of sense […], because they define the training. You can describe a training with it, you can give a bit more to the learner when you have described it. What is the content of the training, what we learn there? What did I have to prepare for? That is the focus. So, such a definition of Learning Goals I find very good."

With respect to the second activity of creating a questionnaire, all four groups liked the efficient subjective measurement of trainees' learning goal achievement on level 2. They agreed that the organisation of questionnaire scales into separate packs in Google Forms helped them to understand the utility of the rubric and supported the reflection on evaluation goals. Especially the groups that considered the evaluation goals of the case were able to select scales accordingly. The groups that did not take the evaluation goals into account struggled in creating the questionnaire. They missed the reasoning for deciding on the selection of scales and investing time of their trainees. The activity stressed the importance of predefined evaluation goals in the design of evaluations and that some partners did not understand the differentiation between the

Kirkpatrick's levels of evaluation. In the same lines, the partners asked for more explanations and a simplification of items that would not be not understandable for trainees.

7.2.4 Pilot 3 – Austria (AT)

We conducted the third pilot with the "Austrian Energy Agency" (AEA), again. This time the micro-training of energy advisors focussed on the topic of management for ecological building materials. The Styrian network once more developed the 2-hour online training as part of their regular voluntary network events. 30 energy advisors participated in the event and training.

As compared to the first pilot, the **learning goal tool was leveraged to guide the re-definition of learning goals for the new training topic**, and in this way **receive feedback on the newly designed artefact**. The evaluation goal remained the same, namely informing the future provision of trainings, suitability of the trainer and relevancy of training contents. AEA still wanted to assess all three levels and implement the same evaluation design.

AEA reconfirmed the results being appropriate for their evaluation goals. The re-design process of the evaluation was perceived as less demanding and time-consuming. In general, the partner reaffirmed the need for explanation for designing an evaluation and selecting corresponding instruments from the rubric. Even though the initial use of the auxiliary learning goal tool was not straightforward, it was seen as making the process much easier and clearer. This was mainly attributed to the tools clarifying the structure of learning goals, providing examples and suggesting verbs. However, it was difficult to select appropriate verbs in correspondence with the level of learning as the list of verbs was too extensive and not yet comprehensive. An overall description of benefits and use of the tool was suggested.

7.2.5 Pilot 4 – Spain (ES)

The fourth pilot was conducted with the Instituto Valenciano de la Edificación (IVE) in the context of testing a certification procedure. This means that there was no actual training of professionals and no chance to check the reactions to learning (level 1). The **newly designed artefacts were used to fictively check the learnings from training (level 2)**, and hence **pilot the reference guide and learning goals tool** as part of the certification endeavour.

IVE assessed, if the knowledge and skills of 15 uncertified practising professionals complies with UNE 85219:2016 standards for window installation. Learning goals were hence inferred from the EU standard from window installation. The test of the certification procedure represented a post-test and included the following elements on level 2:

- Objective measurement –
  - Designed knowledge test via paper-based assessment
  - Observational skill test in an artificial training environment
- Subjective measurement - Self-reported achievement of learning goals (paper-based)

The reference manual could be used to check the certification procedure and assure that the objective measurement with the knowledge test and skill observation followed standards: e.g. application of checklist to guide observation of skill demonstration according to the standardised procedure. For the subjective measurement, the partner appreciated the support in formulating learning goals with the help of the learning goals tool. It was seen as making the process of defining specific and measurable learning goals less time-consuming, but it was required to "learn the usage of the tool". The selection of the level of learning was perceived as complex and action verbs as too focused on cognitive processes and not suitable for skill observation. IVE requested a simplification and explanation of the levels of learning and an adaptation of respective taxonomy

to include more procedural action verbs. They also reaffirmed the benefit of the automatic translation of learning goals for their direct application.

7.2.6 Concluding on EVALUATION123 after Insights from 2 further Pilots and a Consortium Workshop

The **application of the learning goal tool was appreciated in the two pilots and the consortium workshop**. It helped the partners to narrow down their existing general ULOs or unspecified goals into specific measurable learning objectives. The systematic further assured the presence of all elements of a learning goal as suggested by theory (Bloom, 1956; Doran, 1981). This confirms the learning goal tool to be a relevant asset in guiding the design of evaluations and to increase their comparability across interventions. We still had to assist in its application as the six-level taxonomy was too complex and the included action verbs were too extensive and not applicable to the construction and building domain. The adoption could be supported by including automatic translations of learning goals in other languages as not all training designers and trainers are proficient enough in the English language.

The **question packs were also positively perceived in the consortium meeting as they managed to break down the extensive rubric into digestible junks**. This increased the partners' understanding of the utility of the questionnaire scales of the rubric and the three evaluation levels of Kirkpatrick. The partners **appreciated the chance to leverage trainees' self-assessment of learning goals as indication of training effectiveness in limited contexts** that did not allow for objective measures. In line with the implementation of learning goals tools in Google Spreadsheet, question packs were provided in Google Forms and fulfilled the request of a free to use survey offering automatic descriptive analytics. Finally, the reference guide was able to validate the certification procedure in the fourth pilot. **Initial feedback on the reference guide in the form of the digital handbook was positive**. We showcased its role in explaining the usage and integration of the evaluation framework and toolset in the instruction of learning goals tool as part of the pilots and consortium workshop.

Both pilots and consortium meetings underlined that partners have limited resources to design and conduct evaluations. Their **evaluation goals drive their motivation to invest their own and their trainees' resources in an evaluation.** These goals can range from improving a training to providing a certification, for example. However, there is still a **gap in understanding what level of evaluation and which instruments are required for a certain evaluation goal**. Aside providing a step-by-step guidance and toolset for understanding the evaluation framework and rubric, we conclude that the **EVALUATION123 must inform the design of the evaluation as well as the selection of respective instruments in dependence of the evaluation goal**: i.e., stimulate the reflection and guide evaluation design based on the trade-off between evaluation benefits and efforts. Furthermore, we decided to **restructure the question packs** according to the three levels of the evaluation framework (not their stakeholders - training designer, trainer & trainee) and **simplify the taxonomy of the learning goals** tool to ease understanding: 1) Remember & Understand > Understand, 2) Apply > Apply & 3) Analyse, Evaluate, Create > Analyse. The action verbs also needed revision to make them applicable to procedural skills by considering Krathwohl's (2001) revised taxonomy. All of these requirements for the EVALUATION123 lead to **revisions of the reference guide as explanatory manual**: e.g. learning goal taxonomy and learning goal tool or subjective and objective measures in relation to evaluation goals.

The elaboration of the design requirements resulted in an **update and extension of the design propositions and a fourth design artefact** (see Section 3.3 for final EVALUATION123):

1) **Reference Guide (digital handbook/ explanatory manual)**: Explain learning goals in relation to existing high level uniform learning outcomes; explain the use of the learning goals and provide background on underlying taxonomy.
2) **Learning Goals Tool (Google Spreadsheet)**: Simplify the levels of learning and selection of action verbs; adapt taxonomy to suit procedural skills in construction and building domain; provide an automatic translations of learning goals for EU languages.
3) **Question Packs (Google Forms)**: Structure the sets of scales according to the three levels of evaluation (and not the kind of respondent).
4) **Decision Tree (Graphic):** Assist the process of defining the structure of an evaluation based on the first three levels of Kirkpatrick; support making design decisions for evaluation through the analysis and reflection of benefits (intended evaluation goals) and efforts (human, financial and material resources); support the selective and focused reading of the reference guide in dependence of evaluation goal.

7.2.7 Consortium Workshop 2 - NL, FR, AT, IE, ES & SI

We conducted the second co-design workshop on consortium level again with 9 participants from 8 partners: ISSO (NL), Practee Formations (FR), AVE (FR), AEA (AT), IGBC (IE), TUS (IE), IVE (ES) & IRI-UL(SI). The goal of the workshop was twofold once more. It **introduced the newly designed decision tree (see Figure 22) to inform the selection of levels of evaluation as well as corresponding tools** and to answer the question "which decisions have to be made when designing and implementing an evaluation?"

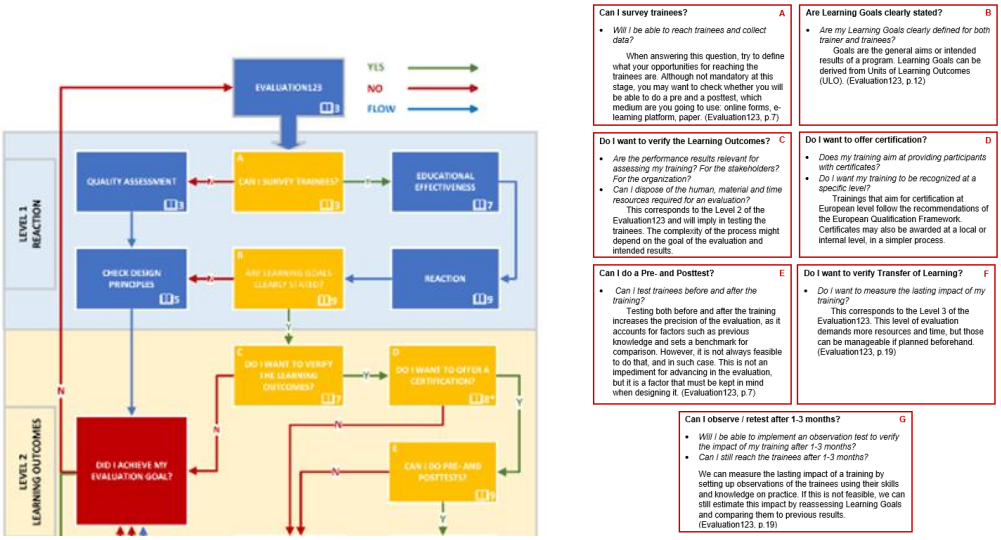


Fig 22: Excerpt from decision tree for reflection on evaluation design and referral to reference guide.

At the same time, the workshop **evaluated the ability of the decision tree to navigate the EVALUATION123 toolset** and trigger reflecting on the trade-off between evaluation benefits and efforts. As the decision tree was acting like a signpost to pages of the manual, the workshop also provided insights on the EVALUATION123 reference guide and toolset. The workshop began with an exploration of experiences in application of the EVALUATION123 since the last workshop. Afterwards, we asked the partners to draft an evaluation plan for a fictive case with the help of the decision tree (see Figure 23) and split them into two groups.

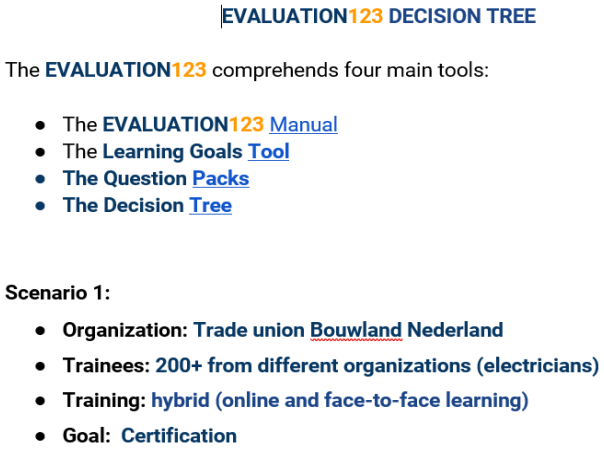


Fig 23: Workshop handout referring to EVALUATION123 and information on the case (scenario).

The consortium could not provide further feedback on the EVALUATION123 reference guide and toolset. No upskilling interventions were conducted in the meantime, except the ones used for pilot 3 & 4. The partners only stressed the importance of simple instructions and efficient evaluations again. Out of this reason, both groups welcomed the support of the decision tree in drafting an evaluation plan even more: "I definitely think this could help in the whole process of designing, training and suitable evaluation because it helps in the steps [of drafting an evaluation plan and] to realise the importance of setting learning goals". The decision tree helped the partners to visualise the evaluation process and to make decisions on their design based on evaluation goals and context: "I like the DT and I think it is a good way to handle the structure [and to] have a map to refer to when you have a complex project as ours". They went through the decision tree multiple times as part of the activity and iterated on the evaluation plan to finally meet their goals. Yet, it was visible that they needed a means to trace the decisions taken. The self-directed usage in both groups indicated the tool being fairly easy to navigate and requiring little explanation. Nevertheless, the activity revealed again the limited understanding of the evaluation framework and its terminology: "it's really nice and clear and simple when you look at it, but when you get into it, you realise it's much more detailed and then you realise that it links to all the other information [and tools in the EVALUATION123]."

7.2.8 Concluding on EVALUATION123 after second Consortium Workshop

In the consortium workshop, we found first evidence for the decision tree triggering reflection on the evaluation design and helping to navigate the EVALUATION123. The **decision tree allowed the partners to explore possibilities and consequences of design decisions for given evaluation goals**. Its simple but central questions disclosed the impact of different decisions in their evaluation plan visually in the diagram. This way, they could 'see and understand' the importance of defining systematic learning goals and the role of the three levels of evaluation for their evaluation goals. In addition, the referrals to the EVALUATION123 allowed them to selectively access additional explanations in the reference guide and decide on the application of the toolset such as the question packs and learning goals tool.

We conclude that the **decision tree motivates users in designing an evaluation according to their needs and makes the EVALUATION123 better accessible**. Nevertheless, the questions, terminology as well as their depiction must be improved to improve on the affordance to iterate the design and link to the rest of the EVALUATION123. This also applies to the explanations in the references guide which need to resemble these improvements. Finally, there is the **need for tracing final evaluation plans**. The defined learning goals and selected instruments allow for discussing as well as improving design decisions for evaluations

with others, and sharing as well as scaling successful and validated evaluation plans and implementation beyond the project.

The elaboration of the design requirements resulted in the **update of the design propositions of the decision tree (see Section 3.3 for final EVALUATION123)**:

1) **Decision Tree (Graphic & Checklist** - see Section 4.1.1**):** Enable tracing evaluation plans including the evaluation context, defined learning goals and selected instruments, for example.

7.3 DETAILED RESULTS OF THE BETA-TESTING (SURVEY STUDY)

Agreement with statements about the reference guide, decision tree, learning goals tool and question packs were assessed on a 6-point rating scale: 1= strong disagreement to 6 = strong agreement. The concrete items, number of participants (N), minimum (Min), maximum (Max), arithmetic mean (M) and standard deviation (SD) of all rating scale items is reported in the following Tables 8, 9, 10, 11, & 12 (with colours indicating the level of agreement). Eleven participants reported their agreement throughout all statements.

	N	Min	Max	M		SD
Q1 It is easy to look up things in the Reference Guide.	11	4	6	5.00	Agree	.894
Q2 The text in the Reference Guide is easy to read.	11	3	5	4.36	Somewhat agree	.674
Q3 The Reference Guide uses too many technical words	11	2	5	3.73	Somewhat agree	1.009
Q4 The images and layout in the Reference Guide facilitate comprehension of the topics.	11	3	6	4.91	Agree	.944
Q5 The Reference Guide provides a good overview of different possibilities for evaluations.	11	3	6	5.09	Agree	.944

Table 8: Descriptive Statistics on usability of the reference guide

	N	Min	Max	M		SD
Q5 The Reference Guide provides a good overview of different possibilities for evaluations.	11	3	6	5.09	Agree	.944
Q6 The Reference Guide helps me understand my evaluation needs.	11	3	6	5.00	Agree	.894
Q7 The Reference Guide helps me understand my evaluation possibilities.	11	3	6	4.82	Agree	1.079
Q8 I feel able to plan an evaluation for a training program using the Reference Guide.	11	3	6	4.91	Agree	.944
Q9 The Reference Guide will be useful for planning evaluations in different contexts.	11	4	6	5.27	Agree	.786

Table 9: Descriptive Statistics on usefulness of the reference guide

	N	Min	Max	M		SD
Q10 The Decision Tree facilitates planning an evaluation.	11	4	6	5.36	Agree	.809
Q11 The Decision Tree helps me aligning my efforts to my intentions for evaluations, getting the maximal evaluation outcomes within my possibilities.	11	4	6	4.90	Agree	.876

Table 10: Descriptive Statistics on usefulness of the decision tree

	N	Min	Max	M		SD
Q12 The Learning Goals Tool helps me specify the ULOs (units of learning outcomes) to units that can be evaluated.	11	1	6	4.09	Somewhat agree	1.375
Q13 The Learning Goals Tool helps me clarifying learning objectives.	11	3	6	4.45	Somewhat agree	.820
Q14 The Learning Goals Tool helps me designing my evaluation.	11	4	6	5.00	Agree	.632

Table 11: Descriptive Statistics on usefulness of the learning goals tool

	N	Min	Max	M		SD
Q15 The Questionnaire Packs help me selecting items for my surveys.	11	3	6	4.73	Agree	.786
Q16 Having the Questionnaire Packs available in digital format (e.g., Google Forms) facilitates their use.	11	5	6	5.55	Strongly agree	.522

Table 12: Descriptive Statistics on usefulness of the learning goals tool

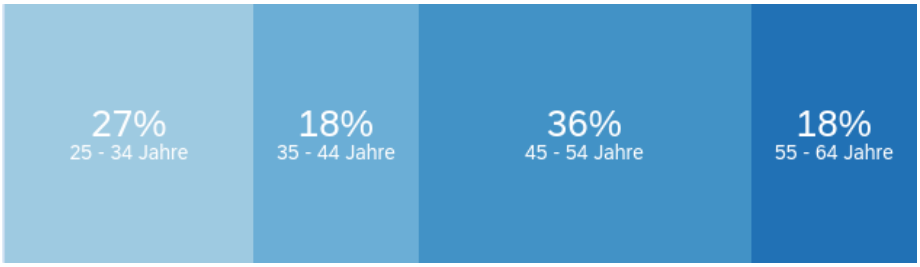
7.4 DETAILED RESULTS OF SELECTED IN-DEPTH EVALUATION – AEA (AT)

Demographics

The demographic data provides information on the profile of the audience targeted by the training. 36.4% of respondents were aged between 45 and 54 years old. The second largest group (27.3%) were aged between 25 and 34 years old. Over half of the respondents (54%) had at least 9 years of experience as energy advisors and 64% have also had an average of 9 years of previous experience in the energy or construction fields.

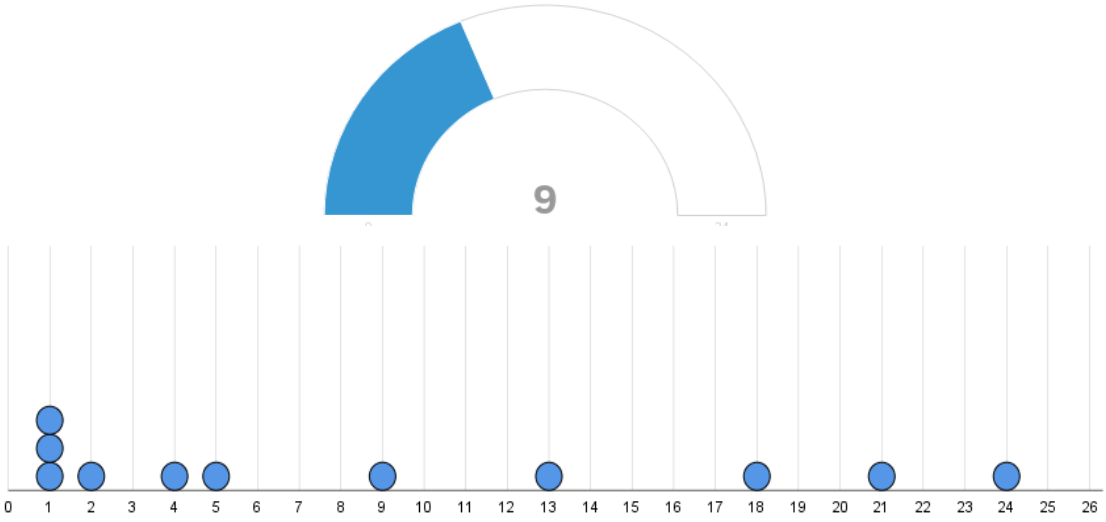
Age Group

Q2.2 – To which age group are you belonging?



Experience as Energy consultant

Q2.5_1 – How many year are you working as energy consultant?



Prior experience in Energy or Construction

Q2.6 - Were you already active in the field of energy transition or in the construction industry before working as an energy consultant?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Yes	7	63.6	63.6	63.6
	2 No	4	36.4	36.4	100.0
	Total	11	100.0	100.0	

Field of experience

Q2.6.1 – If yes, in which role and activity?

- Carpenter, foreman, **site manager,** technician
- **Planning**, Construction **Management**
- **Technicians**, Sales
- Teaching, Scheduling, Purchasing, **Project Management,** Department Management
- **Planner,** Passive House, Sewage Treatment Plants
- Student, Architecture, **Planning**
- **Project manager**

Years of previous experience

Q2.6.2 – If yes, how many years?

#	Field	Min	Max	Mean	Std Deviation	Count
1	Jahre	4.00	10.00	9.00	2.07	7

Learning Goals

The definition of learning goals is at the core of designing instructional programs, as they describe the expected outcomes of the learning experience. For learners, they act as primer for acquiring knowledge and skills, enabling them to assess and modulate their strategies during the learning process.

Five learning goals had been set for the training on *renovation roadmap*, and respondents were asked to self-report their perception of knowledge on the define learning goals before and after the training. However, during the training, only three learning goals were addressed:

- LG1 - describe the most important elements of a gradual renovation.
- LG2 - explain the most important factors for the successful implementation of a renovation roadmap.
- LG5 - independently outline a renovation roadmap.

The data indicated a positive relation between the pre and post training results for all learning goals. Despite the limitations in data analysis imposed by a reduced sample, descriptive results indicated a difference in LG2 and LG5, which were consistent with the ones approached during the training. These resultssupport the relevance of assessing learning goals as means of not only course design and learner's self-regulation, but also serve as an indication for the commissioners of which determined learning objectives have been overlooked or need more attention.

Q3.1 - As part of the short training, various topics were dealt with, which we subsequently formulated as learning objectives. We are interested in your assessment of the extent to which you have achieved these learning goals AFTER the training. Please give your honest assessment of the achievement of the learning objectives in percent:

#	Learning Goal After Training	Mini-mum	Maxi-mum	Mean	Std Deviation	Variance	Count
3.1_5	After the training, I can describe the most important	51.90	100.00	**88.36**	13.81	190.74	11

LG1	elements of a gradual renovation.						
3.1_6 LG2	After the training, I can explain the most important factors for the successful implementation of a renovation roadmap.	59.10	100.00	**87.69**	12.37	153.03	11
3.1_9 LG3	After the training, I can describe the most important processes on a construction site.	2.80	100.00	**66.72**	30.56	934.11	11
3.1_10 LG4	After the training, I can describe the necessary communication interfaces between the trades.	2.80	100.00	**62.16**	31.09	966.40	11
3.1_11 LG5	After the training, I can independently outline a renovation roadmap.	73.90	100.00	**88.94**	8.84	78.14	11

Q3.2 - How much did you already know about the topics of the training or corresponding learning objectives before the training? Please again give your honest assessment of the achievement of the learning objectives BEFORE the training in percent:

#	Learning Goals Before Training	Min	Max	Mean	Std Deviation	Variance	Count
3.2_5 LG1	Before the training, I was able to describe the most important elements of a gradual renovation.	60.00	99.00	**83.27**	11.47	131.65	11
3.2_6 LG2	Before the training, I was able to explain the most important factors for the successful implementation of a renovation roadmap.	53.00	99.00	**77.18**	15.86	251.60	11
3.2_7 LG3	Before the training, I was able to describe the most important processes on a construction site.	3.00	99.00	**62.45**	28.06	787.34	11
3.2_8 LG4	Before the training, I was able to describe the necessary communication interfaces between the trades.	2.00	99.00	**55.91**	29.82	889.36	11
3.2_9 LG5	Before the training, I was able to sketch a renovation roadmap on my own.	63.00	99.00	**78.45**	12.78	163.34	11

Difference in the perception of achievement of Learning Goals

Total Learning Goals	Mean	N	Std. Deviation	Minimum	Maximum
After Training	**78.7745**	11	15.75095	**48.50**	**100.00**
Before Training	**71.4545**	11	17.99135	**42.40**	**99.00**
Mean difference After and Before Training	**7.3200**	11	7.58223	**-6.30**	**18.94**

Satisfaction with the training

Q4.1 - Please indicate your consent to the following statements about the short training as part of the network meeting:

		N	Minimum	Maximum	Mean	Std. Deviation
4.1_1	My decision to participate in this short training was wise.	11	4.00	5.00	**4.8182**	.40452
4.1_2	I was able to learn from the discussions during the training.	11	2.00	5.00	**4.7273**	.90453
4.1_3	I was encouraged to do additional reading or to research the topics discussed.	11	2.00	5.00	**4.1818**	.98165
4.1_4	Based on my experience with this short training, I would like to attend another training in the future.	11	4.00	5.00	**4.9091**	.30151
4.1_5	Overall, the speaker fulfilled my expectations of the short training.	11	2.00	5.00	**4.4545**	.93420

4.1_1	Question	Disagree at all	Partially disagree	Neither	Partially agree	Fully agree	Total
4.1_2	My decision to participate in this short training was wise.	0.00%	0.00%	0.00%	18.18%	81.82%	11
4.1_3	I was able to learn from the discussions during the training.	0.00%	9.09%	0.00%	0.00%	90.91%	11
4.1_4	I was encouraged to do additional reading or to research the topics discussed.	0.00%	9.09%	9.09%	36.36%	45.45%	11
4.1_5	Based on my experience with this short training, I would like to attend another training in the future.	0.00%	0.00%	0.00%	9.09%	90.91%	11
4.1_1	Overall, the speaker fulfilled my expectations of the short training.	0.00%	9.09%	0.00%	27.27%	63.64%	11

Quality of Training

Respondents reported to be extremely satisfied with the ability of the trainer to communicate and transmit the content during the training.

Q4.2 - Please indicate your satisfaction with the short training as part of the network meeting with regard to the following characteristics:

	N	Minimum	Maximum	Mean	Std. Deviation
Quality of the entire training	11	4.00	5.00	**4.4545**	.52223
Enthusiasm of the presenter during the training	11	4.00	5.00	**4.8182**	.40452
Ability of the presenter to communicate the content effectively	11	4.00	5.00	**4.9091**	.30151
Overall Rating	11	4.00	5.00	**4.7273**	.32722

Quality of Interaction

On average, respondents were satisfied with the opportunities for interaction and networking during the training.

Q4.2 - Please indicate your satisfaction with the short training as part of the network meeting with regard to the following characteristics:

Descriptive Statistics				
N	Minimum	Maximum	Mean	Std. Deviation

Opportunity to meet people with similar interests	11	3.00	5.00	**4.1818**	.87386
Interaction with other respondents of the trainings	11	3.00	5.00	**3.8182**	.98165
Opportunity to make new acquaintances in the training	11	2.00	5.00	**3.3636**	1.12006
Overall Rating	11	2.67	5.00	**3.7879**	.88535

Relevance of Training

Respondents were unanimous in stating that this training was relevant for their practice as energy consultants, which confirms the need for the upskilling in renovation roadmaps.

Q5.0 - The contents of the short training and the associated learning objectives are relevant for my work as an energy consultant.

	Frequency	Percent	Valid Percent	Cumulative Percent
Fully agree	11	100.0	100.0	100.0

*100% of the respondents fully agree with the statement.

Questions about the training

Approximately two thirds of respondents claimed to have had all their doubts answered during the training, and 36% (4 respondents) reported still having some type of doubts. Only 3 of those reported what those doubts are. Respondents who reported some type of doubts were also prompted to say what actions they would take to solve them. Most of them reported that they would discuss these topics with colleagues or perform a search for the information they needed.

Q5.1 - Through the training or the subsequent discussion, all my questions about the training content were answered.

	Frequency	Percent
Partially agree	4	36.4
Fully agree	7	63.6
Total	11	100.0

Q5.1.1 - Which of your questions were not answered by this training?

- Which U-values must the components meet after renovation?
- Detailed procedure of a construction site organization with responsibilities

- Ensuring the implementation of meaningful renovations (often is then "saved" in the execution and then a damage is caused or not the optimal out of it - energy consultant is then usually no longer involved

Q5.1.2 - What are you doing to clarify these questions?

Frequencies	Responses N	Responses Percent	% Cases
Q5.1.2_1 Discuss with colleagues	3	33.3%	75.0%
Q5.1.2_4 Search for information (e.g.: Internet search)	3	33.3%	75.0%
Q5.1.2_6 Gain experience through practical application	1	11.1%	25.0%
Q5.1.2_8 Consult training materials (e.g.: PPT slides & notes)	1	11.1%	25.0%
Q5.1.2_7 Other	1	11.1%	25.0%
Total	9	100.0%	225.0%

Motivation

91% of respondents claimed that they feel capable to transfer the newly acquired knowledge/skill into their practice as energy consultants.

Q5.2 - After the training I feel able to apply what I have learned directly in my practice as an energy consultant.

	Frequency	Percent	Valid Percent	Cumulative Percent
Partially agree	1	9.1	9.1	9.1
Fully agree	10	90.9	90.9	100.0
Total	11	100.0	100.0	

Q5.2.1 - I intend to apply what I have learned in the following context. Please describe it briefly.

1. Renovation planning
2. In the course of energy consultations, with a little more detailed knowledge
3. In the consultation
4. Energy consultations
5. Consultations, renovation recommendations, energy certificate improvement texts
6. Consultations on site, inspections with customers on building sites
7. Energy consulting
8. As energy consultant

Troubleshooting

All respondents were asked what kind of action they took whenever they faced a problem related to their practice as energy consultants. Most reported that they would resource to performing a search for the information or would discuss the topic with colleagues more often than contacting NES, training material, instructor or other related professionals.

Q5.1.2 - What do you do if there is a problem?

Response	N	Percent	% Cases
Discuss with colleagues	9	22.5%	81.8%
Search for information (e.g., internet research)	11	27.5%	100.0%
Contact speakers of the short training	3	7.5%	27.3%
Gain experience through practical application	6	15.0%	54.5%
Contact the Network Energy Consulting Styria	6	15.0%	54.5%
Consult training materials (e.B.PPT slides & notes)	5	12.5%	45.5%
Other	0	0.0%	0.0%
Total	40	100.0%	363.6%

All respondents claimed that they would be willing to exchange ideas with their network colleagues to some extent.

Q5.3 - After the short training, I would like to exchange ideas with my network colleagues about the topic and interim experiences.

	Frequency	Percent	Cumulative Percent
Neither nor	6	54.5	54.5
Partially agree	4	36.4	90.9
Fully agree	1	9.1	100.0
Total	11	100.0	

Open Feedback

Q5.4 - Would you also like to give us feedback on the training or do you have any further comments on how to improve it?

1. Great training
2. Very practical and lively explanations of the speaker, thanks again at any time
3. the eternal lyre: online is not the same as on site (especially discussion and exchange of ideas...)
4. Well organized and a very good and competent speaker. Thank you!

5. The topic was very broad and gave a good overview of all fields of action. Maybe in the future trainings on a smaller scale (e.g., what to pay attention to when changing windows, ...), but perhaps a little more precisely.

8 REFERENCES

Anderson, L. W., Krathwohl, D. R, Airasian, P. W., Cruikshank, K. A. & Mayer, R., Pintrich, P. R. Raths, J. & Wittrock. M. C. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Pearson.

Andrade, H. L. (2019). A critical review of research on student self-assessment. In *Frontiers in Education* (p. 87). Frontiers.

Athiyaman, A. (1997). Linking student satisfaction and service quality perceptions: the case of university education. *European journal of marketing*, *31*(7), 528-540.

Berndt, A., Murray, C. M., Kennedy, K., Stanley, M. J., & Gilbert-Hunt, S. (2017). Effectiveness of distance learning strategies for continuing professional development (CPD) for rural allied health practitioners: a systematic review. *BMC medical education*, *17*(1), 1-13.

Bloom, B. S. (Ed.). (1956): Taxonomy of Educational Objectives Handbook 1: Cognitive Domain. New York: Longman, Green & Co.

Bluestone, J., Johnson, P., Fullerton, J., Carr, C., Alderman, J., & BonTempo, J. (2013). Effective in-service training design and delivery: evidence from an integrative literature review. *Human resources for health*, *11*, 1-26.

Burke, L. A., & Hutchins, H. M. (2008). A study of best practices in training transfer and proposed model of transfer. *Human Resource Development Quarterly, 19*(2), 107-128.

Cant, R., & Levett-Jones, T. (2021). Umbrella review: Impact of registered nurses' continuing professional education informed by contemporary reviews of literature. *Nurse Education in Practice*, *50*, 102945.

Cheng, E. W., & Hampson, I. (2008). Transfer of training: A review and new insights. *International journal of management reviews*, *10*(4), 327-341.

Chauhan, R., Ghosh, P., Rai, A., & Kapoor, S. (2017). Improving transfer of training with transfer design: does supervisor support moderate the relationship?. *Journal of Workplace Learning*.

Chao, R. J., & Chen, Y. H. (2009). Evaluation of the criteria and effectiveness of distance e-learning with consistent fuzzy preference relations. *Expert Systems with Applications*, *36*(7), 10657-10662.

Choi, B., Hwang, S., & Lee, S. (2017). What drives construction workers' acceptance of wearable technologies in the workplace?: Indoor localization and wearable health devices for occupational safety and health. Automation in Construction, 84, 31–41.

Cunningham, G. B. (2007). Development of the physical activity class satisfaction questionnaire (PACSQ). *Measurement in Physical Education and Exercise Science*, *11*(3), 161-176.

Doran, G. T. (1981). There's a S.M.A.R.T. way to write management's goals and objectives. *Management Review, 70*(11), 35–36.

Festner, D., & Gruber, H. (2008). Conditions of work environments in fostering transfer of training. In *Emerging perspectives of workplace learning* (pp. 215-231). Brill.

Fessl, A., Maitz, K., Dennerlein, S., & Pammer-Schindler, V. (2021). The Impact of Explicating Learning Goals on Teaching and Learning in Higher Education: Evaluating a Learning Goal Visualization. In *Technology-Enhanced Learning for a Free, Safe, and Sustainable World: 16th European Conference on Technology Enhanced Learning, EC-TEL 2021, Bolzano, Italy, September 20-24, 2021, Proceedings 16* (pp. 1-15). Springer International Publishing.

Ford, J. K., & Weissbein, D. A. (1997). Transfer of training: An updated review and analysis. *Performance improvement quarterly*, *10*(2), 22-41.

Ferguson, L. M. (1998). Writing learning objectives. *Journal for Nurses in Professional Development*, *14*(2), 87-94.

Gao, Y., Gonzalez, V. A., & Yiu, T. W. (2017). Serious games vs. traditional tools in construction safety training: A review. LC3 2017: Volume I – proceedings of the joint conference on computing in construction (JC3), July 4-7, 2017 (pp. 655–662). Heraklion, Greece.

Gao, Y., Gonzalez, V. A., & Yiu, T. W. (2019). The effectiveness of traditional tools and computer-aided technologies for health and safety training in the construction sector: A systematic review. *Computers & Education*, *138*, 101-115.

Garavan, T. N., Morley, M., Gunnigle, P., & McGuire, D. (2002). Human resource development and workplace learning: emerging theoretical perspectives and organisational practices. *Journal of European Industrial Training*, *26*(2/3/4), 60-71.

Gegenfurtner, A., Veermans, K., Festner, D., & Gruber, H. (2009). Integrative literature review: Motivation to transfer training: An integrative literature review. *Human resource development review*, *8*(3), 403-423.

Hattie, J. (2008). Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement (1st ed.). Routledge.

Ho, C. L., & Dzeng, R. J. (2010). Construction safety training via e-Learning: Learning effectiveness and user satisfaction. Computers in Education, 55(2), 858–867.

Kelly, M., Lyng, C., McGrath, M., & Cannon, G. (2009). A multi-method study to determine the effectiveness of, and student attitudes to, online instructional videos for teaching clinical nursing skills. *Nurse education today*, *29*(3), 292-300.

Kirkpatrick, D., 1975. Evaluating Training Programs. Tata McGraw-Hill Education; 1975, Chicago.

Kirkpatrick, D., 2006. Seven keys to unlock the four levels of evaluation. Perform. Improv. 45, 2006.

Kuncel, N. R., Crede, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of educational research*, 75(1), 63–82.

Laker, D. R., & Powell, J. L. (2011). The differences between hard and soft skills and the relative impact on training transfer. *Human Resource Development Quarterly, 22*(1), 111-122. doi:10.1002/hrdq.20063.

Littlejohn, A. (2017) Learning and work: professional learning analytics. In: Lang, C., Siemens, G., Wise, A. and Gašević, D. (eds.) *Handbook of Learning Analytics.* Society for Learning Analytics Research, pp. 269-277.

DeLong, M., Winter, D., & Yackel, C. A. (2005). Student learning objectives and mathematics teaching. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *15*(3), 226-258.

Marzano, R. J. (2010). *Designing & teaching learning goals & objectives*. Solution Tree Press.

McKenney, S., & Reeves, T. C. (2018). *Conducting educational design research*. Routledge.

McNall, M.A in Frey, B. B. (2018). The SAGE Encyclopedia Of Educational Research, Measurement, And Evaluation. *The SAGE Encyclopedia Of Educational Research, Measurement, And Evaluation*, 1187–1190.

Oliver, D., Dobele, T., Greber, M., & Roberts, T. (2004, January). This course has a Bloom Rating of 3.9. In *Proceedings of the Sixth Australasian Conference on Computing Education-Volume 30* (pp. 227-231).

Paavola, S., & Hakkarainen, K. (2005). The knowledge creation metaphor–An emergent epistemological approach to learning. *Science & education*, *14*, 535-557.

Pisklakov, S. (2014). Role of Self-Evaluation and Self-Assessment in Medical Student and Resident Education. British Journal of Education, Society & Behavioural Science, 4(1), 1–9.

Park, Y., Woo, H., Oh, M. R., & Park, S. (2020). The role and related variables of workplace learning in quantitative research. *Industrial and Commercial Training*, *53*(1), 29-59.

Park, S., & Lee, J. Y. (2018). Workplace learning measures for human resource development: review and summary. *Industrial and Commercial Training*.

Pilli, O., & Admiraal, W. F. (2017). Students' learning outcomes in massive open online courses (MOOCs): Some suggestions for course design. *Journal of Higher Education*, *7*(1), 46-71.

Rogaten, J., & Rienties, B. (2021). A Critical Review of Learning Gains Methods and Approaches. *Learning Gain in Higher Education*, *14*, 17-31.

Rolheiser, C., & Ross, J. A. (2001). Student self-evaluation: What research says and what practice shows. Plain Talk about Kids, 43, 57.

Ruiz-Calleja, A., Prieto, L. P., Ley, T., Rodríguez-Triana, M. J., & Dennerlein, S. (2017). Learning analytics for professional and workplace learning: A literature review. In *Data Driven Approaches in Digital Education: 12th European Conference on Technology Enhanced Learning, EC-TEL 2017, Tallinn, Estonia, September 12–15, 2017, Proceedings 12* (pp. 164-178). Springer International Publishing.

Russ-Eft, D. (2002). A typology of training design and work environment factors affecting workplace learning and transfer. *Human Resource Development Review*, *1*(1), 45-65.

Saks, A. M., & Belcourt, M. (2006). An investigation of training activities and transfer of training in organizations. *Human Resource Management, 45*(4), 629-648.

van der Schaaf, M., Donkers, J., Slof, B., Moonen-van Loon, J., van Tartwijk, J., Driessen, E., ... & Ten Cate, O. (2017). Improving workplace-based assessment and feedback by an E-portfolio enhanced with learning analytics. *Educational Technology Research and Development*, *65*, 359-380.

Schiekirka, S., Reinhardt, D., Beibarth, T., Anders, S., Pukrop, T., & Raupach, T. (2013). Estimating learning outcomes from pre-and posttest student self-assessments: a longitudinal study. *Academic Medicine*, *88*(3), 369-375.

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational researcher*, *27*(2), 4-13.

So, H. J., & Brush, T. A. (2008). Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: Relationships and critical factors. *Computers & education*, *51*(1), 318-336.

Tynjälä, P. (2013). Toward a 3-P model of workplace learning: A literature review. *Vocations and learning*, *6*(1), 11-36.

Tzeng, G. H., Chiang, C. H., & Li, C. W. (2007). Evaluating intertwined effects in e-learning programs: A novel hybrid MCDM model based on factor analysis and DEMATEL. *Expert systems with Applications*, *32*(4), 1028-1044.

van Uum, M. S., & Pepin, B. (2022). Students' self-reported learning gains in higher engineering education. *European Journal of Engineering Education*, 1-17.

Wang, Y. S. (2003). Assessment of learner satisfaction with asynchronous electronic learning systems. *Information & management*, *41*(1), 75-86.

Wei, X., Saab, N., & Admiraal, W. (2021). Assessment of cognitive, behavioral, and affective learning outcomes in massive open online courses: A systematic literature review. *Computers & Education*, *163*, 104097.

Xiao, J. (1996). The relationship between organizational factors and the transfer of training in the electronics industry in Shenzhen, China. Human Resource Development Quarterly, 7(1), 55-73.

This project has received funding from the European Union's h2020 framework programme for research and innovation under grant agreement no 892894. The information in this publication does not necessarily represent the view of the European Commission.