>Teacher training for Data Literacy & Computer Science competences

>D4.1 State of the Art Report

train-dl.eu

Prof. Dr. Ulrike Lucke, Martin Reger, Anastasia Tsymboulova | University of Potsdam

Outline

List of tables

List of illustrations

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

Executive Summary

WP4 has the task of evaluation and quality assurance. The policy aspect of the introduction of data literacy (DL) and artificial intelligence (AI) into the educational framework curricula will be investigated. This will be achieved through the evaluation of training courses/teacher trainings/interventions.

The purpose of this State of the Art Report is to provide an overview of relevant scientific literature and results of related projects that can contribute to the further design of the activities in the project. It will additionally provide a brief outlook on the methodology used for the evaluation. Furthermore, the report will be a scientific basis to further inspire the activities in TrainDL.

The report first provides an overview of the basics of political science and some definitions of terms that play a role in policy evaluation (*chapter 1*). Subsequently, the theoretical foundations of policy evaluation are discussed, and the project is classified on the basis of these foundations (*chapter 2*). Afterwards, the current status of the project is briefly presented (*chapter 3*). *Chapter 4* shows the bibliography.

1. Definition of terms

1.1. Data literacy

There is no uniform definition of data literacy (DL) in the literature (see Mandich & Gummer 2013, 30). Often the term is used as a synonym for data science. As work definition we understand: The ability to understand and use data. This implies knowledge of mathematics, statistics, and computer science. It involves identifying, collecting, organizing, summarizing, and interpreting relevant data, as well as formulating hypotheses and problems (see ibid., 30f.). This is missing from current school curricula. Although statistics is taught as part of math classes, DL and the handling of large amounts of data are not. However, the demand for this is increasing as the related issues become more important in science and business as well as in everyday life.

1.2. Artificial intelligence

"Artificial intelligence [...] is a branch of computer science that deals with the automation of intelligent behavior and machine learning." (Lucks 2020) Data and formulas can be used to solve everyday tasks of modern society: from Google searches to vacuum-cleaning robots, Artificial intelligences (AI) achieve superhuman results. They process vast amounts of data in the most abstract and complicated ways to support daily decisions. The research field of AI is concerned with the application of algorithms to this end, which go beyond just predetermined calculations and attempt, among other things, to mimic human learning, perception, and action in lines of code.

1.3. Statecraft

In German language, there is the term "Politik", which is derived from the Greek word "polis". This refers to statecraft (since "politics" as a English term has a different meaning like shown in the next sub-chapter). It refers to the statecraft by which societal problems are counteracted and regulated; for example, through resolutions, measures, enforcement of demands and goals, implementation of programs, and distribution of resources (e.g., tax money) (see Sager et al. 2021a, 2).

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

1.4. Policy, polity and politics

A distinction is made between the (1) structural dimension of order (polity), (2) the procedural dimension of power (politics), and (3) the substantive/content-wise dimension of design (policy) (see Heidenheimer 1986, 4; Sager et al. 2021a, 2). According to this model, the terms are defined as follows:

(1) *Policy* refers to the substantive aspect of statecraft: for example, something that a government or party intends to introduce or change through reform. A policy always refers to specific policy areas (see Rohe 1994, 61f.; Schmidt 2013, 207). Within the TrainDL project, a policy proposal for the sub-area of education policy is being developed. This is a planned reform about the extension of the framework curricula by the (compulsory) topics DL and AI. A policy is, among other things, about the formulation of tasks and goals as well as the conception of political programs. Social contents, values and interests are reflected in this policy dimension (see Rohe 1996, 6f.). In the evaluation of TrainDL, the content level is specifically examined.

(2) In contrast, the *polity* dimension deals with the political framework for action, the conditions under which statecraft generally takes place and the state constitution - more precisely: legal order, separation of powers and basic forms of organization (see ibid., 64f.).

(3) *Politics* encompasses the conflict of power relations in the selection of personnel to implement policies (see ibid., 62).

1.5. Policy cycle

The policy cycle is an ideal-typical sequence of the policy process according to Jann & Wegrich (2014). Figure 1 shows the individual phases.

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

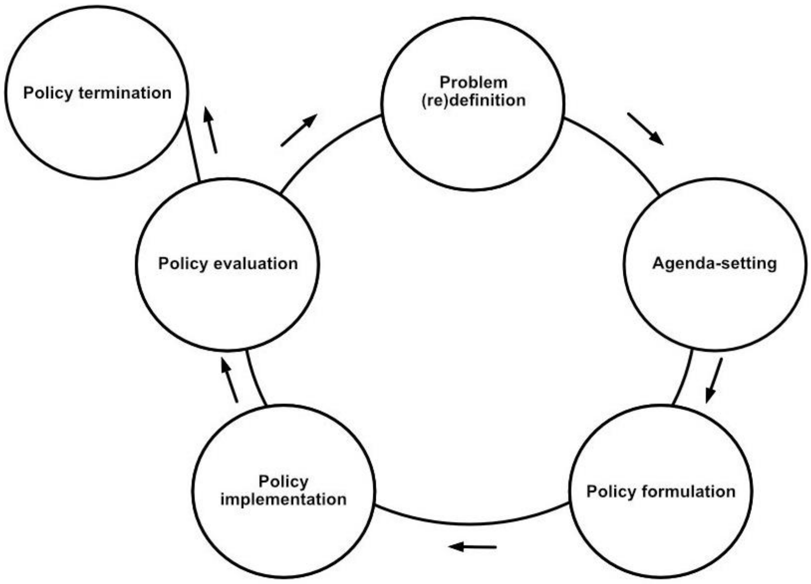

***Figure 1**: The ideal-typical policy cycle (Jann & Wegrich 2014, 106)*

Problem definition involves the recognition of a societal-social problem with a need for control (see Jann & Wegrich 2014, 107). In this context, this corresponds to the introduction of DL and AI into teacher training and curricular frameworks and thus as compulsory topics in school education. Then, in the *agenda-setting* step, this articulated problem is put on the political agenda. Specifically, the agenda is perceived either by the mass media, the professional public, or politically by the government and parliament (see ibid., 107f.). This was explicitly achieved through the decision to implement the TrainDL project and to allocate resources to it. The *policy formulation* step is about formulating alternative courses of action and goals up to concrete laws, budgets, regulations, government statements, reforms or policy recommendations. It is TrainDL's job to develop recommendations for action. Characterized by debates on the part of the public and all other stakeholders – such as competent administrations, working groups or interest groups – the concrete formulation is a multifaceted process (see ibid., 120). Decision-makers (such as government, parliament) make the decision at the end of this debate, subject to all interests, opinions, scientific knowledge as well as ideological orientations. After the concrete policy has been formulated, *policy implementation* of decisions takes place. This means that measures are taken, resources and

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

services are distributed, norms are applied, contracts are signed and decisions are implemented (see ibid., 120): For example, the framework curricula for schools change and DL and AI are now integrated into computer science classes, or the relevant content is added to teacher training programs. As a rule, the changes are made by the responsible ministries and authorities (see ibid., 110). The *policy evaluation* step examines whether the formulated policy has had its desired effect or whether previously defined goals have been achieved and what impact it has had. Evaluation research is a subfield of policy research and represents an important part of the policy process (see ibid., 120). In the context of TrainDL, an attempt is made to anticipate this step through multiple policy experimentation evaluations (see chapters 2. Policy eveluation and 3. Current status of the evaluation of TrainDL). On the basis of this, a decision is made as to whether a *policy reformulation* will take place and the policy cycle will start all over again, or whether a *policy termination* will be made. The TrainDL project is part of the policy formulation process, as it is intended to contribute to producing decision-making tools for the resolutions needed. The goal is to formulate/publish recommendations based on a (final) evaluation. However, since it is part of the project to evaluate trainings, in the following we will only talk about *policy evaluation*.

2. Policy evaluation

Policy evaluation is the scientific and empirical assessment of policy design, implementation and effectiveness (see Sager et al. 2021a, 2). Such evaluations are about the empirical identification of causal effects through an intervention: "Evaluation is assessment" (ibid., 3). The TrainDL project is not about evaluating a policy, but about simulating an evaluation (of the policy) using a policy experimentation approach. Actually, the project is in the policy formulation stage in the policy cycle, as there is no instruction(s) yet on how to train teachers about DL and AI. However, in the course of the experiment, policy proposals, mediated via or in the form of training (in three iterative rounds of interventions), are evaluated.

Thus, in the following, we will speak exclusively of policy evaluation and not and policy formulation, since for the purpose of the evaluation it is simulated that these training concepts are implemented policies in order to apply the methods and approaches of policy evaluation.

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

The specific process and methodology are briefly described in chapter 3. To provide an overview of the entire scope of policy evaluation, Figure 2 shows what will be discussed in the next subsections and how adjustments have been made or are planned for implementation in the TrainDL project. Clockwise from top to bottom, the ovals of the mind map correspond to the content of the next subchapters.

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

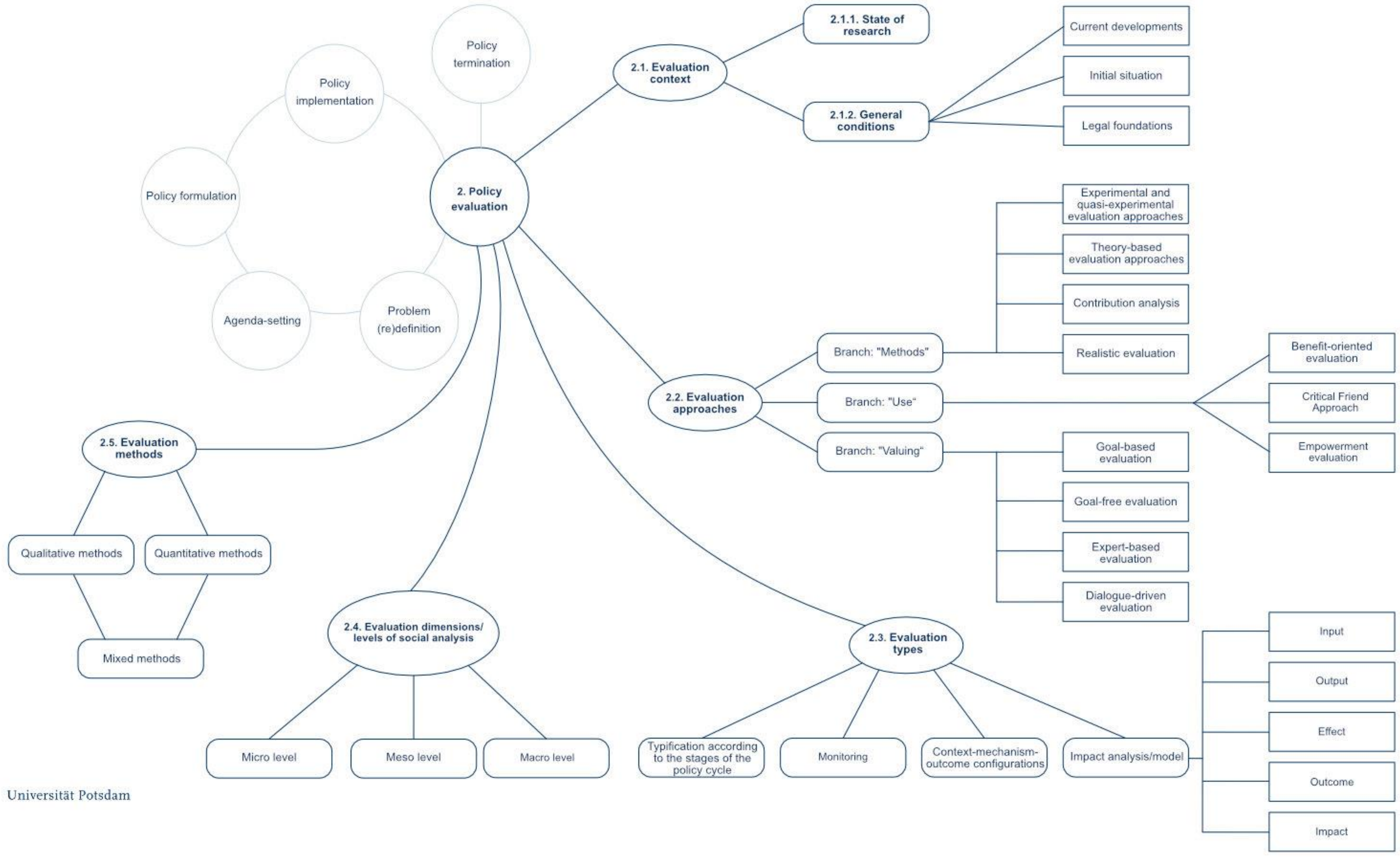

Figure 2: Policy evaluation overview

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

In the following, the chapters highlighted in bold with theirs corresponding contents are presented.

2.1. Evaluation context

2.1.1. State of research

In preparation for the evaluation, the evaluation context is considered. The state of research and the general conditions (initial situation of the social problem, current developments and legal basis; see also following chapter 2.1.2.) form the context for the study. In the following, the related research is presented first, which at the same time anticipates the current developments from the general conditions.

DL and AI are gaining importance on societal, informatics, economic, and political levels as well as in education in recent decades. Data from all areas of science and business are generated in unimaginable quantities every day. According to the International Data Corporation (IDC), the global volume of data will increase fivefold from 33 zettabytes in 2018 to 175 zettabytes by 2025 (see Reinsel, Gantz, & Rydning 2018, 3). This makes data science (and the associated data literacy) an emerging science with methods from computer science, mathematics and other areas of science, depending on the respective data nature. Thus, the demand for data scientists is also increasing, as is the need to introduce the topics of DL and AI into school education so that early awareness of these multidisciplinary topics can take place. Figure 3 shows the increase in scientific literature since 2010, with research on AI in education more than doubling within the last decade.

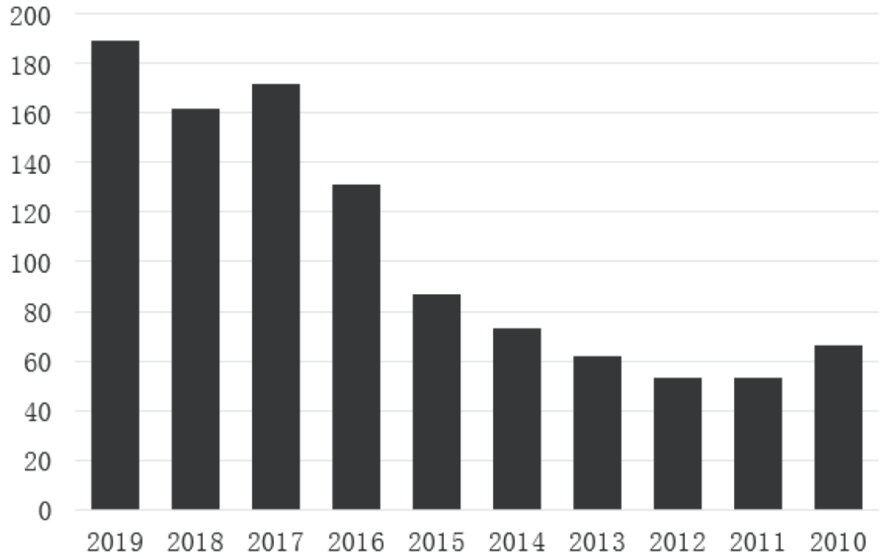

Figure 3*: Papers in Web of Science and Google Scholar of the last 10 years (Chen et al., 2020)*

The literature on integrating DL and AI into students' education distinguishes two aspects: (1) using both topics to support teaching and make it more efficient, and (2) actually introducing the content into the framework curricula so that students are taught it. TrainDL focuses on the second aspect. DL and AI are already taught worldwide, but almost never at the school level: Until 2022, with few exceptions, both topics were only established in university framework curricula (see Wu et al. 2022, 1f.). The topics are taught for application in academic work or in the context of it, but not for student teachers to teach to students. First pilot projects at lower educational level have been conducted e.g. in the US and Germany (see Ali et al. 2019, Heinemann et al. 2018; Pedró 2019). In Table 1, a brief exemplary overview of selected DL and AI education projects for students is presented to illustrate the state of research.

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

Source	Description of the project
„Constructionism, Ethics, and Creativity: Developing Primary and Middle School Artificial Intelligence Education“ (Ali, Payne, Williams, Park & Breazeal 2019)	The experiment by Ali et al. (2019) involves three lessons for 225 students from Cambridge (UK) – from 5th to 8th grade – with no prior knowledge of DL and AI. In respective 45-minute sessions, the focus was on basic workings of AI, an initial stimulation of creative computational thinking, and a focus on the ethics behind it. The curriculum developed is intended to provide an introduction to AIs, supervised machine learning, and algorithmic biases. Google's Teachable Machine (Creative Labs Google 2022) is used to teach a classification problem, and Cathy O'Neil's Ethical Matrix is used for decision and stakeholder analysis (O'Neil & Gunn 2020). Next, they developed their own algorithm that decides how to make the best peanut butter and jelly sandwich, and finally, the students revise the YouTube suggestion algorithm. The children are confronted with the ethical challenges behind each decision and there is extensive discussion. Surveys in the form of worksheets completed before and after the intervention show that the children can reflect on and have understood learned content, but to varying degrees.
„Data Science and Big Data in Upper Secondary Schools: What Should Be Discussed From a Perspective of Computer Science Education?“ (Heinemann et al. 2018)	Heinemann et al. (2018) designed a pilot curriculum for 19 students from Paderborn, Germany, in upper secondary grades 11 and 12. The curriculum was three hours a week for a full year. Content was divided into four modules: (1) First, basic statistical methods around big data with the goal of promoting DL awareness. (2) Then AI with a focus on machine learning with classical algorithmic data-driven problem solving techniques and an introduction to programming with Python. Decision Trees and Artifical Neural Networks were created. (3) After this module, students worked on big data in their own projects. (4)The last module should encourage the young people to reflect on what they have learned as well as on the social, cultural aspects, opportunities and risks, and the role of humans as data scientists.

Table 1: Overview of selected projects regarding the application of DL & AI in the classroom

In the experimental setup of the TrainDL project, no content about DL and AI is taught to students, but rather it is primarily investigated how teachers can be efficiently further trained and educated in order to then teach this content. However, there is no scientific literature, evidence or studies on this. TrainDL is the first time a training unit on DL and AI is evaluated on a scientific basis. Therefore, it is only mentioned where this content has already been taught to teachers in pilot projects (but these have not been scientifically studied). An UNESCO paper on 2019 reports on individual projects worldwide. Here, the range of content is from initial programming skills to the teaching and application of AI in the classroom. Without citing other sources, it reports the development of new curricula in the EU, UK, Estonia, Argentina, Singapore, and Malaysia. In France, South Korea, and China, plans are being developed to prepare for a world with AI by strengthening the education sector (Pedró 2019, 6).

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

2.1.2. General conditions

TrainDL – "Teacher training for Data Literacy & Computer Science competences" – takes place within the Erasmus+ program of the European Union. The program promotes Europe-wide cooperation in all educational sectors as well as youth and sports. Within three iterative field test rounds within a policy experimentation, educational concepts for DL and competences for AI will be developed. Its central goal is "to provide evidence-based recommendations for the structural implementation of data literacy and AI skills in curricula and education systems across Europe" (Gesellschaft für Informatik e.V., 2022).

2.2. Evaluation approaches

A policy evaluation needs one or more approaches. These serve to set priorities within the evaluation and at the same time form the basis of the approach. In this chapter, an overview of different approaches is given. Figure 4 illustrates the so-called tree model of (Alkin & Christie 2004), oriented to the content of Sager, Hadorn, Balthasar & Mavrot's overview of selected evaluation approaches (2021b). Evaluation in the context of TrainDL is guided by the framework, set by other WPs, and the opportunities presented by the interventions. No approach has been specifically chosen, but rather it is a combination of several approaches. The most likely overlaps are with the realistic evaluation and the Critical Friend Approach.

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

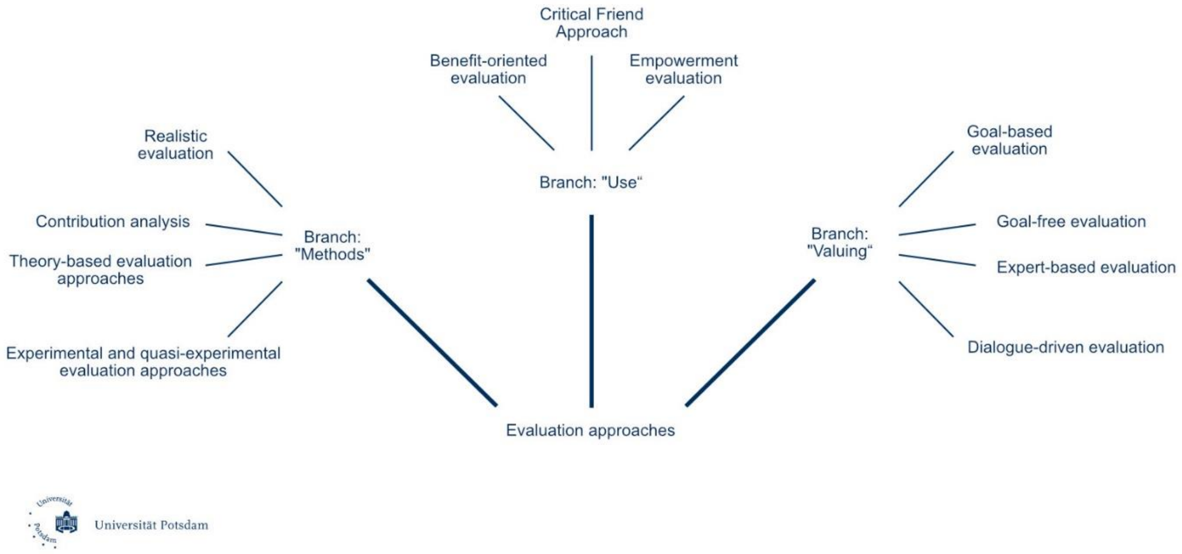

***Figure 4**: Evaluation approaches according to Sager et al. 2021b*

The following Table 2 provides an overview of the focus of these selected approaches and classifies the TrainDL project in these contexts.

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

***Table 2**: Evaluation approaches according to Balthasar (2021b)*

Branch of evaluation		Evaluation approaches		Reference to TrainDL
Branch: "Methods"	Goal: Clearly attribute effects to intervention	Experimental and quasi-experimental	Some of the founders of the approach are Donald T. Campbell and Julian C. Stanley. By assigning experimental and control groups, the aim is to ensure that effects were induced only by the intervention. In pure experimental approaches, this is ensured by random sampling; in quasi-experimental approaches, it is ensured by clever selection of comparable groups of participants. This approach cannot explain why an intervention works or not, only that the effects are clearly attributable to it.	The TrainDL project is neither an experimental nor a quasi-experimental approach. There is no control group, only a participant group.
		Theory-based	Huey-Tsyh Chen and Peter Rossi founded the "theory-driven evaluation". This places the evaluation focus on the "why." The functionalities, concepts and structures behind the intervention are used to explain why the intervention works. This has the advantage of identifying why goals are achieved or not achieved, but lacks a scientific basis.	The impact model of TrainDL is presented in chapter 2.3.2. In-depth: impact models for evaluation.
		Contribution analysis	John Mayne designed the analysis approach in which the focus is on the contribution, the change through the intervention. In a process-oriented manner, an impact model is designed, evidence on the effects is collected, alternative explanations are searched for, the impact model is further developed, new evidence is sought, and finally the model is improved again.	The TrainDL project follows a similar process. The framework conditions of the individual trainings (length, content, target group, etc.) are repeatedly adjusted between the intervention phases, but not the impact model. Nevertheless, new findings are collected in several runs, which contribute to an overall assessment.

Branch: "Methods"	Goal: Clearly attribute effects to intervention	Realistic evaluation	Ray Pawson and Nick Tilley developed an approach that focuses on external contextual factors. They argue that these are the critical elements that trigger the impact of the intervention. Context-mechanism-outcome configurations (see also chapter 2.3.1. Overview) are modele das follows: - Context = geographic, socioeconomic status quo, values, and norms, - Mechanism = functioning and effectiveness of the intervention, - Outcomes = changes over time In addition, there are empirical tests that are designed to test hypotheses about the effect relationships. An advantage of the approach is that alternative explanatory contexts are included and the risk of erroneous results is minimized. A disadvantage may be that the evaluation could be costly.	The TrainDL project is strongly oriented towards the realistic evaluation. Through the planned before and after survey, the outcome level (see chapter 2.3.2. In-depth: impact models for evaluation) will be evaluated. In the electronic surveys the contextual factors (e.g. number of years of education, attitude towards DL and AI etc.) will be collected and also detailed by the face-to-face interviews. In the same way, the mechanism is attempted to be elicited through the competency/self assessment test (see chapter 3. Current status of the evaluation of TrainDL). Here, one mechanism would be that optimally designed training is likely to increase the knowledge of the teachers. Thereupon it could be that they incorporate the contents in their instructions. Likewise, previously established hypotheses will be empirically tested.
Branch "Use"	Goal: Usefulness of evaluation studies, involve stakeholders in the evaluation process	Benefit-oriented evaluation	Michael Patton developed an approach that focuses on the users, i.e., those affected by the changes. They should be involved in the evaluation from the beginning. First, the individuals and groups are identified, then the expectations for the study are worked out with those affected, then data is collected and finally processed. The interactive process has the advantage of increasing the likelihood that the results will actually find use. A potential disadvantage is that there might be a lack of willingness to cooperate on the part of the target group.	The evaluation directly involves the affected target group of teachers. Student teachers and (former) teachers helped to prestest questionnaires and interviews. The target group is interviewed through questionnaires before and after the interventions. Also in the oral interviews some of the teachers have the possibility to influence later interventions. Pupils are not included as a target group to be interviewed.

Branch "Use"	Goal: Usefulness of evaluation studies, involve stakeholders in the evaluation process.	Critical Friend	Andreas Balthasar's approach focused on the benefits for those responsible for the programme. It is about providing targeted and direct support to those responsible, as a critical friend would do. As an active, external advisor, the evaluators point out the need for action. On the other hand, objectives and evaluation design are strictly methodological. One advantage is that the evaluators are particularly familiar with the topic due to their advisory role. A disadvantage can be that they are not independent or that the independent view is distorted from the outside.	The evaluators from WP4 take a critical advisory role and pursue the task of improving the interventions through evaluation by elaborating action points. Programme managers are the other WPs (WP2, WP3) who design and lead the trainings.
		Empowerment evaluation	David M. Fettermann has developed another participatory approach by helping programme managers, staff and beneficiaries to reflect on the intervention through evaluation. This is done through external consultation and observation. Participants in the intervention are asked to help conduct the evaluation, which makes it more likely that problem-solving strategies can be found. The participation of all stakeholders increases the chances of success and stimulates discourse between the groups. A lack of willingness to cooperate can have a negative effect. In addition, the increasing number of stakeholders can increase the time and money costs.	Only the colleagues of WP2, WP3 and WP4 as well as the participants of the training are involved in the planned evaluation.

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

Branch "Valuing"	Goal: evaluation of the effect is in the foreground	Goal-based evaluation	Goal-based evaluations check whether goals have been achieved and whether the "target" state corresponds to the "actual" state. This is done by empirically testing the defined goals and then the results are compared with the goals. One weakness of this approach is that sometimes goals can only be formulated in an abstract or blurred way, which makes testing more difficult. Furthermore, the approach does not provide sufficient why an intervention did not work.	The competence test assesses whether the training has achieved the goal of educating teachers in DL and AI.
		Goal-free evaluation	Michael Scriven designed the goal-free approach, in which the aim is to disregard the objective of the intervention as much as possible so that the evaluation can capture the actual, unbiased effect on those affected. This has the advantage that non-intended effects are taken into account and possible changes in the objectives hardly play a role. Nevertheless, the disadvantage remains that an evaluation is mainly concerned with whether explicitly formulated goals have been achieved.	The evaluation tries to capture impacts outside the objectives through the questions in the questionnaires and interviews, but still focuses on the achievement of the objectives (education in DL and AI, formulations of recommendations based on this).
		Expert-based evaluation	J Blaine Worthen and James Sanders formulated the approach of expert-based evaluation. Experts from the same field (but at a different organisational level) should carry out the evaluation. In the process, this expertise is drawn upon to make use of the relevant professional knowledge. This has a negative impact on the non-intended effects, as the focus remains on the topic bubble.	In WP4, a mixture of internal and external (multidisciplinary trained) evaluators is active. This way, the view of 'external' and 'internal' expertise is maintained at the same time.

Branch "Valuing"	Goal: evaluation of the effect is in the foreground	Dialogue-driven evaluation	Egon G. Guba and Yvonna S. Lincoln designed the approach with the aim of actively increasing the participation of stakeholders in the evaluation so that their interests can be incorporated into the negotiation process. In doing so, the different value systems of stakeholders in the conflict of interests are addressed during the evaluation. One disadvantage is the challenge of communication.	There is communication between WPs within regular meetings, but not actively with all stakeholders.

After defining an evaluation approach, the concrete evaluation model/type for the intervention can be designed on the basis of an evaluation type (see Bussmann et al. 1997).

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

2.3. Evaluation types

2.3.1. Overview

Evaluation types are ideal types of an evaluation which can be used as a model or scheme for orientation for the concrete design. Depending on the evaluation approach and focus, some types are more or less appropriate (see ibid.). Table 3 presents some selected evaluation types and compares them with the TrainDL evaluation model/concept.

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

Evaluation types		Reference to TrainDL
Typification according to the stages of the policy cycle	Busmann, Klöti, & Knoepfel (1997) have developed a typification of evaluations based on the policy cycle (see chapters 1.5. Policy cycle and 2. Policy evaluation). Their approach is practice-oriented and at the same time close to political science. In doing so, the authors modified the policy cycle to the following eight stages: Policy concept, Policy design/administrative programme, authority management for policy implementation, action plans for execution, outputs, impacts, outcomes, policy evaluation results. These levels form the evaluation objects with respective criteria. The focus can be on one or more objects (see ibid., 69ff.).	In the evaluation concept of TrainDL, the terms output, impact and outcome play a big role. The exact definitions are explained in more detail in Table 4, respectively in the next chapter.
Impact/analysis model	In an impact model, assumptions are made about the expected interrelationships between - the goals, the measures and structures for implementation (*input*), - the achievements, what political actors do (*output*), - the effects on the target groups, why they do it (*outcome*) and - the effects on those affected, what they achieve (*impact*) (Sager & Hinterleitner 2014, 109ff.). There are different subtypes of impact models with varying definitions of input, output, outcome and impact.	Input, output, outcome and impact were defined as categories of analysis. On the one hand, to be found in the Figure 5 and defined in the Table 4 (see next chapter), where the different interpretations of the terms by several authors are compared. Table x then also presents the definitions used in TrainDL.
Context-Mechanism-Outcome-Configurations	To map the interactions between context and public policy, so-called context-mechanism-outcome configurations (CMO configurations) are modelled. This means that in specific social, cultural, local, historical and institutional contexts (C), certain mechanisms (M) triggered by a public policy become effective and lead to corresponding outcomes (O). Within the framework of the evaluation, it is now necessary to analyse which CMO configurations are present in each case and to what extent a public policy has an effect (Befani et al. 2007, Sager and Andereggen 2012, Sager and Hinterleitner 2014).	One of the outcomes measured in this evaluation are competency/self assessment tests (see chapter 3. Current status of the evaluation of TrainDL). These can be modified by the mechanism of the training content and considered within the context (of the intervention).
Monitoring	According to Sager & Hinterleitner (2014), monitoring is not an evaluation, but a "routine, permanent and systematic collection of comparable data" (ibid., 439) to determine changes in the behaviour of target groups caused by an intervention. In contrast to an evaluation, explanations are not collected to understand the effectiveness of an intervention. According to Balthasar et al. (2021b), however, this type of data collection can be classified in the branch "Methods" (see chapter 2.2.) of evaluation approaches.	Through the three evaluation phases, the TrainDL project operates a kind of monitoring. Systematic comparable data is collected to capture the changes in impacts related to the changes in the training/target groups.

Table 3: *Evaluation types (according to Bussmann, Klöti and Knoepfel 1997)*

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

For better understanding, the terms related to impact, etc., should be explained, as there are different conceptions depending on the author.

2.3.2. In-depth: impact models for evaluation

Models are a simplified representation of reality. They help to work out the basic interrelationships of complex issues. In the context of policy evaluation, models are applied to arrive at findings that shape recommendations for action as well as to verify to what extent the policy solves the societal problem or not. In the following, Table 4 gives a brief overview of which impact models/policy evaluation models exist. In the same table, at the end, TrainDL's model is also defined.

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

***Table 4**: Impact models/policy evaluation models in comparison*

Source	Input	Output	Effect	Outcome	Impact
Impact model according to Bussmann, Klöti and Knoepfel (1997)	The authors do not include an input as an analytical category in their impact model. Nevertheless, their definition of the policy concept serves as a basis for later input formulations. The first stage of policy generation concludes with a policy concept, which is to regulate the social problem by the state.	Output here refers to all **direct services** provided by governmental and non-governmental actors in the course of a policy - for example, training services, taxes, subsidies, controls, prohibitions, permits, etc. For evaluation purposes, data on temporal, spatial and addressee-specific distributions should be collected and evaluated together in terms of impact and outcome data.	The authors do not include an effect as an analytical category in their impact model.	All "intended and unintended, desired and undesired, **direct and indirect effects**" (110, own emphasis) are attributable to outcome. This includes any changes in behavior, living conditions, or other effects on the environment or society that are attributable to the policy.	The "real effects of public policies on policy addressees" (p. 103). This is about the **extent of behavioral changes** after policy implementation. At the same time, the aim is not "to capture the actual behaviors, but [...] impact relationships between a policy and its outputs and the perceived behaviors of its addressees."
Impact model by Befani, Ledermann and Sager (2007) and Ledermann et al. (2006)	In Sager and Ledermann (2007) as **policy concept** and in Ledermann et al. (2006) both **policy concept and input**. This includes a "definition of the political problem and the state's options for action, as well as the totality of legal provisions and instructions for a political program" (Ledermann et al. 2006, 4).	Labeled as **performance** in Sager and Ledermann (2004) and named as both **output and performance** in Ledermann et al. (2006). This includes the sum of all end products of the policy process.	The authors do not include an effect as an analytical category in their impact model.	The outcome represents the most significant impact on the addressees in the form of **behavioral changes** and is evaluated on the basis of effectiveness and impact-related efficiency.	Impact is the second most important effect on the beneficiaries, whose **living conditions** are expected to improve as a result of the policy change. This is also evaluated on the basis of effectiveness and impact-related efficiency.

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

Linear impact model of Sager and Hinterleitner (2014, 444), after Bussmann, Klöti and Knoepfel (1997, 70)	Again, the input is **the policy concept** that is supposed to solve the social problem.	**Services** under the policy- change to bring about the outcome and behavioral changes.	The authors do not include an effect as an analytical category in their impact model.	**Behavioral changes** directly caused by the policy.	Impact here, as in Bussmann et al. (1997), refers to the **change in society** that a measure has caused. It evaluates whether the situation for the beneficiaries has improved as planned in the policy concept/input.
Schröder and Kettiger (2001)	The authors do not include an input as an analytical category in their impact model.	Here, too, output refers to the **services** provided, but specifically starting from administrations and in relation to their addressees or customers.	Effect is defined by Schröder and Kettiger (2001) as demonstrable, **immediate effects**, in the sense of improvements brought about by the policy.	Outcome is the **indirect effect** of the policy. In contrast to the effect, it goes beyond the direct effect and includes consequential effects.	The **subjective effect** on the recipient is the impact. It is evaluated by taking into account the needs and backgrounds of the recipients. Accordingly, it goes beyond the objective output and evaluates it based on the context and implementation.
Impact levels of social work interventions by Uebelhart and Zängl (2015)	The authors do not include an input as an analytical category in their impact model.	Output is defined differently here than in the previously presented literature. It refers here to **reaching the target** group and whether the "activities" (73) took place as planned.	The effect refers to the **demonstrable impact** of the policy. In the model, this corresponds to the acceptance of the policy offer by the target group(s).	The outcome here is the **indirect effect** of the policy on the addressees.	Impact represents the **subjective effect** on the addressees after policy implementation.

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

TrainDL: Project level	The policy to be evaluated is based on an **evaluation concept** developed by WP4. This includes the framework of the training (content, duration, format, length, etc.). After the final evaluation, the matured concepts can be published as proposals.	This refers to the services provided within the framework of the TrainDL project, the **three intervention rounds**, the knowledge conveyed therein and the materials provided. Because no 'real' or existing policy is being evaluated – but a policy within a study – no 'real' administrative performance can be evaluated.	The demonstrable effect is the **increase in knowledge** of the participants. This is measured by competency tests (see chapter 3. Current status of the evaluation of TrainDL). In addition, other questions are compared and **assessments and opinions** are given and evaluated as part of the data analysis. Furthermore, questions about the training will be used to assess its effectiveness.	This concerns **behavioral changes and immediate effects**; the latter can be recorded, for example, by measuring the participants' increase in knowledge - the extent to which they are well enough equipped to teach DL and AI. A planned follow-up survey will assess the extent to which content on DL and AI has been implemented or what barriers to this have been identified.	The change in living conditions/contexts and society also resulting from the outcome should be a **broad/committed teaching of DL and AI**. Pupils should be exposed to DL and AI at an early age and be enabled to understand and consider them as a future career choice. This should also contribute to combating the shortage of skilled workers in the relevant field.

Together with the levels of social analysis to be presented in the next chapter, these perspectives allow us to better classify target levels of evaluation questions.

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

2.4. Evaluation dimensions / Levels of social analysis

In (social) science, the perspectives of the micro, meso and macro levels are used to identify the extent of the problems, the effects and the people affected: A social problem is often not attributable to just one dimension, but to several or all of them. In order to maintain an overview and not to lose any perspective, these classifications are made. (1) The micro level describes individuals or their social actions (this includes communication – referring to Krotz 2008, 44). (2) The meso level examines social entities – e.g. organizations, informal groups like family or the social institution marriage. (3) The macro level examines society or its fields of action/subsystems such as the educational system (see Esser 1996, 112, Fleige 2011, 48).

The following model results from a combination of the explanations from the evaluation dimensions (levels of social analysis – this chapter) and the effects of the TrainDL project (application level), combining the dimensions to the objects of investigation. This results in the following figure 5.

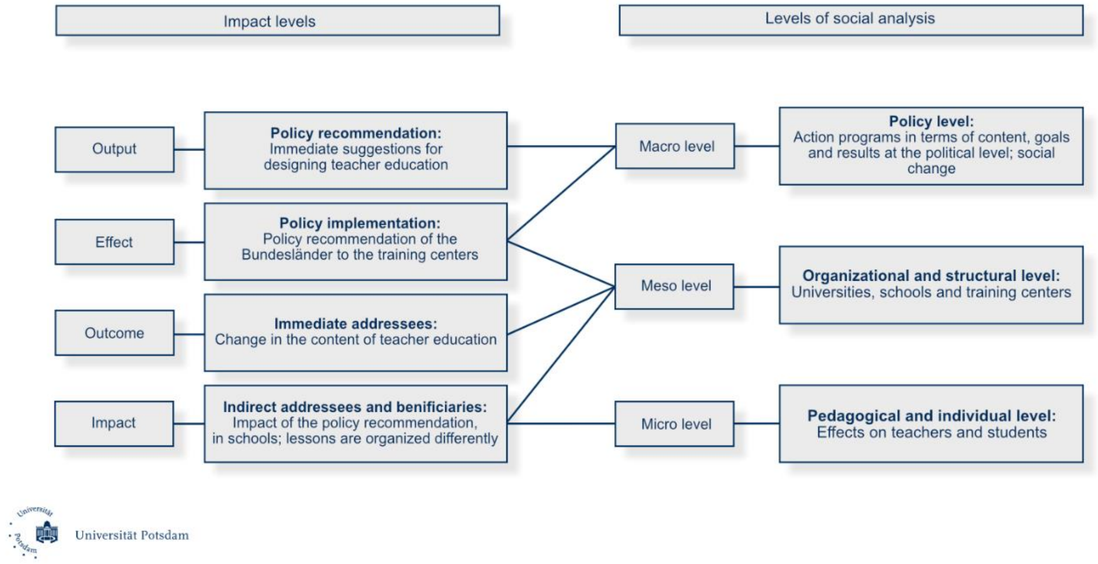
Impact levels
Levels of social analysis
Output
Policy recommendation: Immediate suggestions for designing teacher education
Macro level
Policy level: Action programs in terms of content, goals and results at the political level; social change
Effect
Policy implementation: Policy recommendation of the Bundesländer to the training centers
Meso level
Organizational and structural level: Universities, schools and training centers
Outcome
Immediate addressees: Change in the content of teacher education
Impact
Indirect addressees and benificiaries: Impact of the policy recommendation, in schools; lessons are organized differently
Micro level
Pedagogical and individual level: Effects on teachers and students
Universität Potsdam

Figure 5: *Combined model of impact levels and levels of social anaysisis*

At the *micro level*, the actors are teachers, student teachers and students who would be affected by the policy change. On the *meso level*, the actors are the institutions in which teaching takes place, i.e., further education/training centers that have to modify/expand their course offerings and the chairs in the universities that will adapt their module offerings and the schools in which teaching takes place differently. In terms of the *macro level*, it is about reflecting on overall societal implications due to a 'new' generation of students building up the first contact with DL and AI while still in school. This includes initial instruction on the societal implications of these topics and the basic tenets of their scientificity. This increases the likelihood that some of the students will decide to choose a career in this direction. Thus, the growing need for data scientists is met.

2.5. Evaluation methods

Research methodology comprises the procedures and analytical techniques used to clarify the research question(s) in a study. These are divided into quantitative, qualitative, and mixed-methods approaches (see Döring & Bortz, 1984). Here, quantitative methods refer to the collection of numerical measurements from samples using standardized measurement instruments to test hypotheses (see ibid., 23). Qualitative methods, on the other hand, involve the collection and interpretive analysis of verbal, visual, and/or audiovisual data in a "deliberately nonstructured manner to few cases" (ibid., 25). A mixture of both methods is used in the evaluation of the TrainDL project to produce maximum insight (referring to ibid., 27). The combination of methods can be conducted sequentially or simultaneously. An ideal typical sequence of both methods (quantitative and qualitative) can be seen in the following Figure 6 in which usual steps in 9 phases of the research process from problem formulation to data presentation are shown; in qualitative methods, this also includes partially circular processes.

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

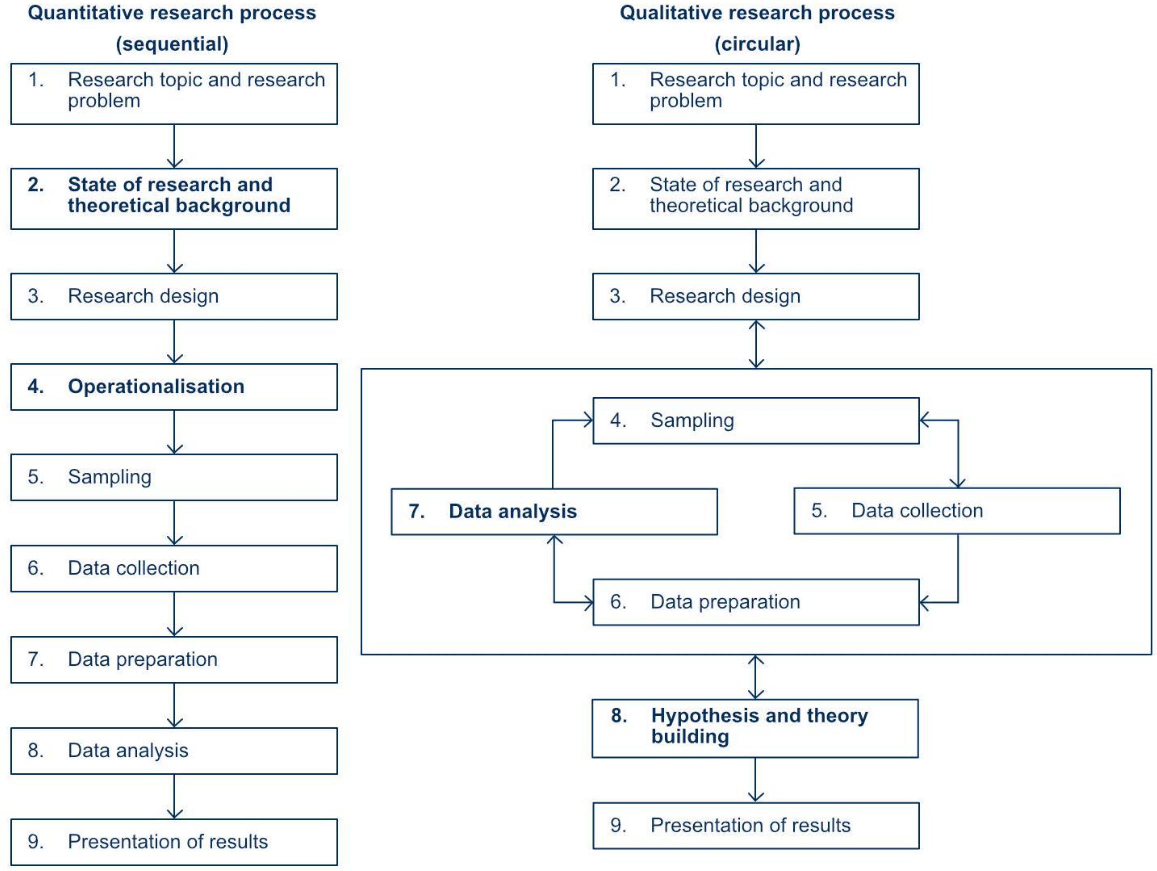
Quantitative research process (sequential)
1. Research topic and research problem
2. State of research and theoretical background
3. Research design
4. Operationalisation
5. Sampling
6. Data collection
7. Data preparation
8. Data analysis
9. Presentation of results
Qualitative research process (circular)
1. Research topic and research problem
2. State of research and theoretical background
3. Research design
4. Sampling
5. Data collection
6. Data preparation
7. Data analysis
8. Hypothesis and theory building
9. Presentation of results

Figure 6: *Schematic representation of the quantitative and qualitative research process, content according to Döring & Bortz (1984) (design/colur scheme slightly altered due to English translation)*

The evaluation process in the TrainDL project will be discussed more in the next chapter.

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

3. Current status of the evaluation of TrainDL

According to the project plan, the evaluation of the interventions in the framework of TrainDL should concern the methodological, pedagogical, organizational and political perspective of the trainings. Figure 7 shows an excerpt of the project application, which classifies these aspects.

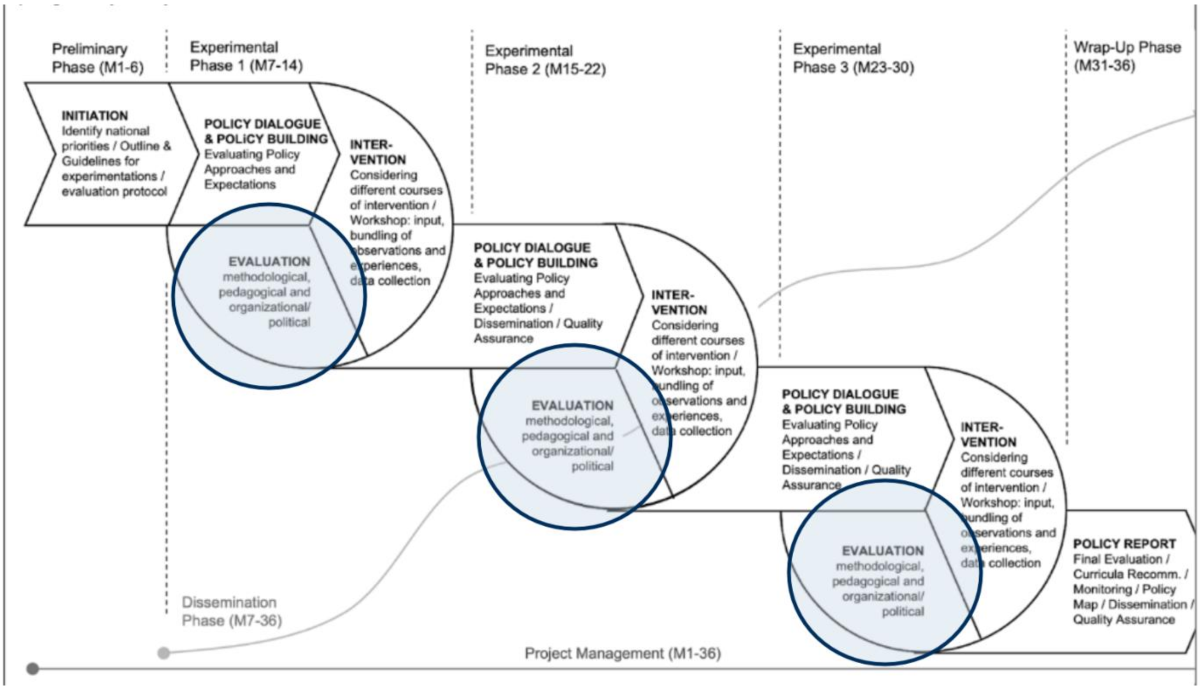
Preliminary Phase (M1-6)
Experimental Phase 1 (M7-14)
Experimental Phase 2 (M15-22)
Experimental Phase 3 (M23-30)
Wrap-Up Phase (M31-36)
INITIATION
Identify national priorities / Outline & Guidelines for experimentations / evaluation protocol
POLICY DIALOGUE & POLICY BUILDING
Evaluating Policy Approaches and Expectations
INTER-VENTION
Considering different courses of intervention / Workshop: input, bundling of observations and experiences, data collection
EVALUATION
methodological, pedagogical and organizational/ political
POLICY DIALOGUE & POLICY BUILDING
Evaluating Policy Approaches and Expectations / Dissemination / Quality Assurance
INTER-VENTION
Considering different courses of intervention / Workshop: input, bundling of observations and experiences, data collection
EVALUATION
methodological, pedagogical and organizational/ political
POLICY DIALOGUE & POLICY BUILDING
Evaluating Policy Approaches and Expectations / Dissemination / Quality Assurance
INTER-VENTION
Considering different courses of intervention / Workshop: input, bundling of observations and experiences, data collection
EVALUATION
methodological, pedagogical and organizational/ political
POLICY REPORT
Final Evaluation / Curricula Recomm. / Monitoring / Policy Map / Dissemination / Quality Assurance
Dissemination Phase (M7-36)
Project Management (M1-36)

Figure 7*: TrainDL project design (Project proposal, 39)*

We are currently in the survey and evaluation phase of the first intervention round. Data has been collected and is being processed and analyzed. Which data is collected and how is explained in more detail below.

A mixed-methods approach will be used to quantitatively measure – with (electronic) questionnaires and competency/self assessment tests before and after the training to measures before-after effects, among other things. In addition, qualitative guideline interviews will be conducted immediately after the intervention, and follow-up surveys will be conducted approximately six months later at school. This task organisation is shown below in Figure 8.

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

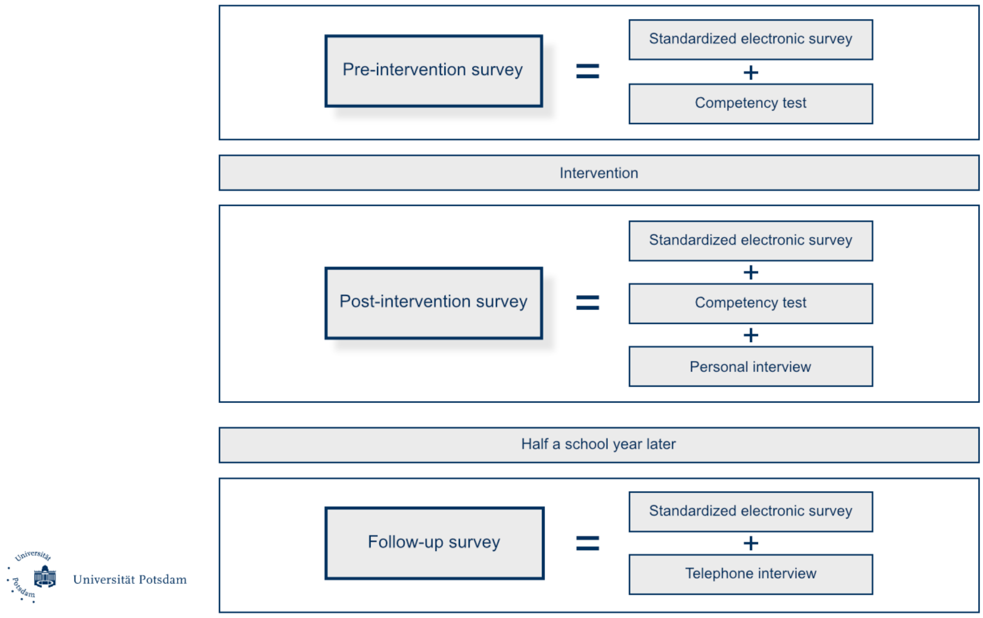

***Figure 8**: Procedure of an intervention survey*

These tasks are generally intended to be repeated for each intervention, with limitations based on timeline or specific formats/target groups.

The following Table 5 shows roughly which data are to be collected with the respective methods at which points in time.

***Table 5:** Surveys and their goals and contents*

Surveys	Goals and contents
Pre-intervention survey: Standardized electronic survey (questionnaire)	- Questions about personal/demographic data (e.g., age, gender, stage in apprenticeship training, work experience, etc.) - Questions on DL and AI attitudes and usage patterns; some of these questions will be repeated both before and after training to measure direct effects
Pre-intervention survey: Standardized electronic survey (competency/self assessment test)	- Knowledge about DL and AI before training with selected competency issues as well as self-assessment questions for comparison with results after training: to measure the direct effect due to the intervention
Post-intervention survey: Standardized electronic survey (questionnaire)	- Attitude after training and estimated applicability as well as possible usage behavior of DL and AI; also in comparison with the pre-survey - Questions on the assessment of the suitability of the selection of topics and examples of use/exercises presented in the intervention

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

Post-intervention survey: Standardized electronic survey (competency/ self assessment test)	- Same questions as in the pre-survey; to measure direct effects due to the intervention
Post-intervention survey: Personal interview	Personal interview with a part of the participants concerning topics that are linked to all evaluation dimensions/levels (chapter 2.4.) of social analysis (micro, meso, macro level), among others: - Opportunity to talk about problems and challenges - Reasons for attending intervention, expectations, questions about level of difficulty - If and what kind of content about DL and AI has already been taught, experiences with it or reasons why no integration of DL and AI has been done so far - Views regarding integration of DL and AI in teacher education and framework curricula - Question about institutional obstacles, barriers - Question about possible changes on the societal level - Feedback on intervention
Follow-up survey: Standardized electronic survey, Telephone interview	- Verification whether the (planned) change in usage behavior has taken place/was possible. - Success, failure in introduction of DL and AI or reasons for non-introduction - Record changes in behavior

The methods and their application are described in detail in Deliverable D4.3 Description of the Evaluation Methodology.

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

4. List of references

Ali, S., Payne, B., Williams, R., Park, H. W., & Breazeal, C. (2019). *Constructionism, Ethics, and Creativity: Developing Primary and Middle School Artificial Intelligence Education.* MIT Media Lab Cambridge.

Alkin, M. C., & Christie, C. A. (2004). *An Evaluation Theory Tree.*

Befani, B., Ledermann, S., & Sager, F. (2007). Realistic Evaluation and QCA: Conceptual Parallels and an Empirical Application. *Evaluation*, 25-46.

Bussmann, W., Klöti, U., & Knoepfel, P. (1997). *Einführung in die Politikevaluation.* Helbig und Lichtenhahn.

Chen, L., Chen, P. & Lin, Z. (2020). Artificial Intelligence in Education: A Review. IEEE Access, Vol. 8, 75264-75278, 2020.

Creative Labs Google (27.06.2022). *Teachable Machine*. https://teachablemachine.withgoogle.com/

Döring, N., & Bortz, J. (1984). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften.* Springer.

Esser, H. (1996). *Soziologie. Allgemeine Grundlagen*. 2nd ed. Campus.

Fleige, M. (2011). *Lernkulturen in der öffentlichen Erwachsenenbildung. Theorieentwickelnde und empirische Betrachtungen am Beispiel evangelischer Träger*. Waxmann.

Fröhlich, R. (2005). Zur Problematik der PR-Definition(en). In R. Fröhlich, P. Szyszka, & G. Bentele, *Handbuch der Public Relations - Wissenschaftliche Grundlagen und berufliches Handeln* (103-120). Springer VS.

Gesellschaft für Informatik e.V. (02.07.2022). *TrainDL. About the Project.* https://train-dl.eu/en/about-traindl/project

Heidenheimer, A. (1986). Politics, Policy and Policey as Concepts in English and Continental Languages: An Attempt to Explain Divergences. *The Review of Politics*, 3-30.

Heinemann, B., Budde, L., Schulte, C., Rolf, B., Frischmeier, D., Poworny, S., & Wassong, T. (2018). Data Science and Big Data in Upper Secondary Schools: What Should Be Discussed From a Perspective of Computer Science Education? *KIT Scientific Publishing*, 1-18.

Jann, W., & Wegrich, K. (2014). Phasenmodelle und Politikprozesse: Der Policy-Cycle. In K. Schubert, & N. C. Bandelow, *Lehrbuch der Politikfeldanalyse* (97-132). De Gruyter Oldenbourg.

Krotz, F. (2008). Kultureller und gesellschaftlicher Wandel im Kontext des Wandels von Medien und Kommunikation. In T. Thomas, *Medienkultur und soziales Handeln* (43-62). VS Verlag.

Ledermann, S., Hammer, S., Sager, F., Dubas, D., Rüefli, C., Schmidt, N., Trageser, J., Vettori, A. & Zeyen Bernasconi, P. (2006). *Evaluation der Strategie "Migration und Gesundheit 2002–2006". Schlussbericht und Beilagenbände 1-3*. Bundesamt für Gesundheit.

Lucks, K. (2020). *Der Wettlauf um die Digitalisierung. Potenziale und Hürden in Industrie, Gesellschaft und Verwaltung.* Schaeffer-Poeschel.

Mandich, E., & Gummer, E. (2013). A Systemic View of Implementing Data Literacy in Educator Preperation. *Educational Researcher*, 30-37.

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

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. (2006). A Proposal for the Darthmouth Summer Reaserch Project on Artificial Intelligence. *AI Magazine*, 12-14.

O´Neil, C., & Gunn, H. (2020). Near-Term Artificial Intelligence and the Ethical Matrix. In M. Liao, *Ethics of Artificial Intelligence* (237-270). Oxford University Press.

Pedró, F. (2019). *Artificial Intelligence in Education: Challenges and Opportunities for Susteinable Development.* UNESCO Education Sector.

Reinsel, D., Gantz, J., & Rydning, J. (2018). *The Digitization of the World - From Edge to Core.* IDC White Paper.

Rohe, K. (1996). Politische Kultur: Zum Verständnis eines theoretischen Konzepts. In O. Niedermayer, & K. von Beyme, *Politische Kultur in Ost- und Westdeutschland* (1-21). Springer Fachmedien Wiesbaden.

Sager, F., & Andereggen, C. (2012). Dealing with Complex Causality in Realist Synthesis: The Promise of Qualitative Comparative Analysis (QCA). *American Journal of Evaluation*, 60-78.

Sager, F. & Hinterleitner, M. (2014). Evaluation. In K. Schubert & N. Bandelow, *Lehrbuch der Politikfeldanalyse.* 3rd ed. (437-462). De Gruyter Oldenbourg.

Sager, F., Hadorn, S., Balthasar, A., & Mavrot, C. (2021a). Begriffliche Grundlagen. In F. Sager, S. Hadorn, A. Balthasar, & C. Mavrot, *Politikevaluation. Eine Einführung* (1-15). Springer VS.

Sager, F., Hadorn, S., Balthasar, A., & Mavrot, C. (2021b). Überblick über ausgewählte Evaluationsansätze. In F. Sager, S. Hadorn, A. Balthasar, & C. Mavrot, *Politikevaluation Eine Einführung* (65-96). Springer VS.

Schmidt, M. G. (1997). Vergleichende Policy-Forschung. In D. Berg-Schlosser, F. , Müller-Rommel, *Vergleichende Politikwissenschaft. Uni-Taschenbücher, Vol 1391* (207- 221). VS Verlag für Sozialwissenschaften.

Schröder, J. & Kettiger, D. (2001). *Wirkungsorientierte Steuerung in der sozialen Arbeit. Ergebnisse einer internationalen Recherche in den USA, den Niederlanden und der Schweiz. Band 229. Schriftenreihe des Bundesministeriums für Familie, Senioren, Frauen und Jugend*. Kohlhammer.

Uebelhart, B., & Zängl, P. (2015). Social Policy Making. In B. Wüthrich, J. Amtstutz, & A. Fritze, *Soziale Versorgung zukünftig gestalten* (65-88). Springer VS.

Wu, D., Xu, H. S., & Lv, S. (2022). What should we teach? A human-centered data science graduate curriculum model design for iField schools. *Jasist Wiley*, 1-19.

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