Deep-learning API-feature Specification Tool for Formative Assessment in Workshops

Wassim Mahfouz
Faculty of Computer Science and Automation
Ilmenau University of Technology, Germany
wassim.mahfouz@tu-ilmenau.de

Heinz-Dietrich Wuttke
Faculty of Computer Science and Automation,
Ilmenau University of Technology, Germany
dieter.wuttke@tu-ilmenau.de

***Abstract*- This Research to Practice Full Paper is driven by the question: In workshops to develop deep-learning API-features, how can trainer be effectively supported in specifying and formative assessing of individual tasks (i.e. the API-features) for each student? In formative assessment, the trainer's main difficulty in specifying the tasks is to ensure that the challenges in a task are calibrated to the particular needs of a student at a particular time (i.e. the student's available skills). While goals that are very high challenging lead to a student's anxiety, goals that are too easy generate student's disengagement and boredom. This paper presents a specification tool designed to support the trainers in practicing Hattie's Visible Learning [1] pedagogy in dialogues for tasks specification and formative assessments. Its design enables two support functions in visualization; 1) a function to declare visually a task specification to make its mastery goals clear and progressively challenging, and 2) a function to refactor a task specification for visualizing formative assessment feedback in dialogues for reflection. Besides the two functions, its innovative design enables the flow model based analysis to understand student engagement. The analysis is according to Csikszentmihalyi's flow model to diagnose, in dialogues for formative assessments, one of student's eight emotional states (i.e. boredom, apathy, worry, anxiety, control, arousal, and flow) in terms of challenge level and skill level. To verify the tool effectiveness in trainer-student Freirean dialogues, we conducted our experimental tests in workshops where students with different culture backgrounds, and different levels of prior skills have tasks to develop API-features based on deep-learning principles and algorithms. The algorithms selected in the test cases are for automatic classification of images or of text documents. The test results show the effective use of the tool to create conditions to overcome student's negative emotions (e.g. worry or anxiety), and help him/her experience positive emotions (e.g. control or flow) in learning and mastering the algorithms; to put it differently, to create situations in initial specification and co-specification of tasks, where student's skills are engaged by progressively higher challenges to understand the deep-learning based classifier algorithms.**

Keywords*- *formative assessment in engineering education, deep-learning, Hattie's visible learning pedagogy, Freire's dialogic pedagogy, Csikszentmihalyi's flow model.

I. Introduction: Literature Review and Main Contributions.

Assessment is usually understood to have two purposes: summative and formative. Summative assessment is used to judge students' achievements at the end of an instructional unit, whereas formative assessment aims to gather evidence about students' learning progress to influence teaching methods and priorities [2]. According to Hattie [3], using two dividing notions (i.e. formative and summative) is not helpful in education debate about assessment. To improve the debate Hattie suggests to use, instead of two dividing notions, the single notion "Assessment-capable learner". This notion enables Hattie to argue that the main key of formative assessment is to teach the students how to evaluate their own learning [3]. Driven by the question "What works **best** for learning" Hattie reports the outcomes of his research in "Visible Learning" [1, 3]. Visible Learning means an enhanced role for teachers as they become evaluators of their own teaching. According to Hattie, Visible Learning and Teaching occurs when teachers see learning through the eyes of students and help them become their own teachers. Hattie claims that the notion "Assessment-capable learner" is what the education debate needs because it communicates and highlights the importance of both formative and summative assessments. To support his claim he used the metaphor suggested by Robert Stake [4] that draws a clear distinction line between formative and summative assessments and highlights the importance of them both. The metaphor is: "*When the cook tastes the soup, that's formative and when the guests taste the soup, that's summative*". That means: for the task of a teacher as evaluator and the task of the student as self-evaluator; both evaluating paths during learning (i.e. formative assessments) and evaluating a goal achievement in a learning path (summative assessments) are helpful and important. In stake's metaphor, tasting the soup by cook during making it and by guests during serving it is helpful and important. The major argument presented in "Visible Learning" [1] is that when teaching and learning are visible, there is a greater likelihood of students reaching higher levels of achievement. To make teaching and learning visible requires an accomplished teacher, as evaluator and activator, who knows a range of learning strategies to build the students' surface knowledge, deep knowledge, and conceptual understanding. The teacher needs to provide direction and redirection in terms of the content being understood, and thus make the most of the power of feedback [5]. The teacher also needs to have the skill to get out of the way when learning is taking place and the student is making progress towards meeting the criteria against which successful learning will be judged. In their book "Turning Point for the Teaching Profession" [6] Rickards, Hattie et al are arguing for the establishment of teaching as a true "profession" alongside areas such as medicine or law. Moreover, in their book, they are elaborating on evaluative thinking and clinical practice as the basis of this new profession. Hattie's Visible Learning pedagogy was developed in context of primary schooling, therefor our investigations in his pedagogy, during our deductive based research activities, aim to test if his pedagogy and its explaining constructs hold or do not hold in contexts of our university offers for freshman students or for different culture based student groups. The offers are for teaching and training software design and modern programing languages. In context of these offers, understanding the pedagogy-assessment linkage is the main objective of our research as software architects. In our research based architecting activities, we formulate our research questions (e.g. questions in [7, 8]) to practice two types of research; namely deductive and inductive. In deductive research we start from one or more theories in pedagogy and/or psychology (e.g. Hattie's Visible Learning pedagogy or Freire's pedagogy of the oppressed [9]), whereas in inductive research we start from phenomena observations. The educational philosophies of John Dewey [10], Paulo Freire [11], and other advocates of progressive and critical pedagogy can be regarded, according to Noam Chomsky [12], as further developments of Humbold's philosophy for educational reforms started in Germany about 200 years ago. In terms of educational contexts, present-day german universities may seem far removed from the challenges and concerns that troubled Freire when he began to develop his ideas of liberatory education and pedagogy of the oppressed. However, we are interested in investigating if Freire's dialogic pedagogy and its explaining constructs hold or do not hold in context of some offers provided by us for freshman students, and for students with immigration background because of the wars in Europe (e.g. war in Ukraine) and in Middle East (e.g. war in Syria). Our research based architecting activity in [7] was about the **F**ormative **A**ssessment (hereafter **FA**) implementation challenges in context of project based pedagogy for students in third semester of three years based computer science program, whereas in [8] the research based activity was about the FA implementation challenges in course based pedagogy context. The context was for a teacher of Computer Organization (CO) course that has about 130 freshman students. In contrast to [7, and 8], in this paper we are interested in investigating and understanding how to integrate dialogue practices (like Freire's dialogic pedagogy), to offer **responsive** FA actions in our workshops. The notion "action" is a modelling notion conceptualized by us in [7] to enable modeling a set of FA actions in our research questions. For example in [7] we have formulated our question as it follows: *Within limited time resources available to trainers in projects for Big Data Analytics (BDA) problems, how can they define project requirements for FA actions?*. The answer to the last question was a suggested BDA APIs architecture designed as helping tool for trainers. It helps them clarifying visually student challenges. In [7] clarifying visually student challenges is labeled by us as our main FA action because it reflects technically our initial interpretations of Hattie's Visible Learning theory in context of our projects based pedagogy. These technical interpretations work as input arguments to enable testing if Hattie's Visible Learning theory and its explaining constructs hold or do not in our projects based pedagogy context. Beside input arguments to enable testing Hattie's Visible Learning theory, the architecture in [7] provides input arguments to enable testing not only Csikszentmihalyi's flow theory [13, 14] but also Buckingham Shum and Deakin Crick's theoretical framework for dispositional learning analytics [15]. In contrast to the project based context in [7] and the course based context in [8], the research based architecting activity in this paper is about the workshop based pedagogy context and its FA implementation challenges/problems**.** To put it differently, given student diversities in education backgrounds, in cultural backgrounds, and in prior skills, communicating goals and learning intentions in dialogue practices for tasks specification and formative assessments are not only challenging but also time consuming. To face this difficulty, this paper suggests a specification tool designed initially to facilitate practicing Hattie's Visible Learning pedagogy. Hattie's pedagogy focuses on making learning and teaching visible in classroom settings for precollege students, whereas this paper focuses on integrating and testing his pedagogy in our university workshops for freshman students. Beside the design requirements to test, integrate and modify Hattie's pedagogy to fit the workshops settings, the paper shows how the tool design is developed to fit an existing set of our tools and practices. The practices aim to integrate Freire's dialogic pedagogy, and Csikszentmihalyi's flow theory in synthesizing our arguments and theoretical views. In context of our existing tools and practices, the specification tool boundary and the metrics for measuring its effectiveness (see section 5.a for more details) are defined to include two key support functions as solution for two FA challenges/ requirements in trainer-student Freirean dialogues organized in the workshops. The requirements are:

Requirement 1: trainer can produce a visual report in goals setting and clarifying**.** Via visual declaration of an API-feature, the goals for learning and mastery must be clear and progressively challenging. Mastery goal is one type of the self-goals that student can have. There is a rich literature on the goals that students can have. According to Hattie [1], there are three types of goals; mastery, performance, and social. The theoretical views for mastery goal [1] and self-concept [16] as introduced by Hattie are what we plan to focus on and use as guidelines in our goal specification practices .

*Requirement 2**:*** trainer can produce a report to visualize assessment feedback. Via visual refactoring of an API-feature to facilitate dialogue based formative assessments, the assessment feedback must be visible for reflection and action.

Finally, the last requirements drive our contributions in this paper. The main contributions are:

- We have designed a support tool for specifying deep-learning API features (sections 2, 3 and 4). Its innovative design enables two support functions for visualization in declaring and refactoring the API-features. The functions can be effectively used by the trainer to produce and individualize visual reports for two purposes:
    1. To clarify goals for learning and mastering algorithms. The visual reports clarify goals with progressively higher challenges to understand deep-learning principles and algorithms.
    2. To visualize assessment feedback in dialogue practices. The assessment feedback visualized by the reports enables reflections for action. This maximizes clarity in goals communication to better understand the deep-learning algorithms.
- To understand student engagement, we present a practice method for integrating the flow model based engagement analysis, Hattie's pedagogy and Freirean dialogue based formative assessments (sections 2 and 3).

To sum up, we conduct our research based architecting activities to find out and design the best support tools for helping teachers and trainers solve social-technical problems in responsive practices. In next section we present the new specification tool in the big picture of our dialogue based responsive practices.

II. Our Specification Tool in Big Picture

Fig. 1 shows in big picture how our specification tool is designed to facilitate the best responsive practices in tasks specification. The tasks are about understanding deep-learning principles and algorithms. The figure highlights also how the tool enables responsive practices in two contexts: macro and micro. The macro context is about practices for evaluative thinking in workshops management, whereas the micro context is about dialogue practices for formative assessments (e.g. Freirean [9] dialogues) organized during workshops. In macro context, the main management task is the refinement of our previous practices to plan next improvement actions, whereas in micro context, the refined practices are put in action at micro levels. By micro levels, we mean the implementation challenges of the responsive practice at hand. For these implementation challenges, we run workshops in iterations because each iteration enables us to refine our decisions for improvements.

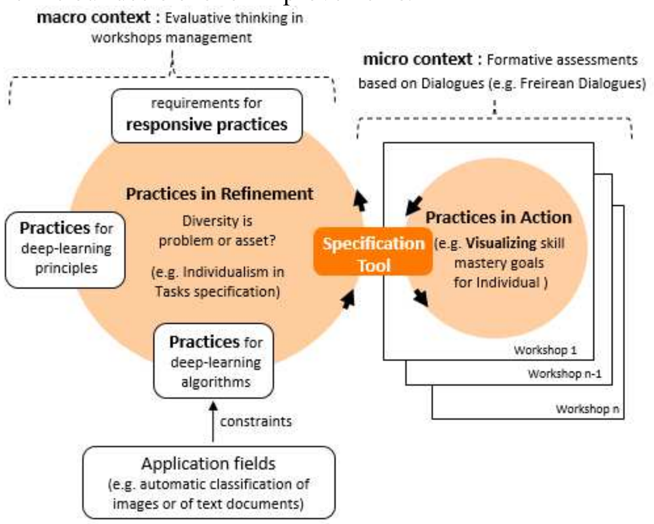

Fig. 1 our specification tool in big picture: It facilitates best responsive practices in tasks specification.

To give a concrete example in context of practicing Hattie's pedagogy, table (1) documents a practice for the refinement task: *evaluate if students' diversity is a problem or an asset, and based on the evaluation outcomes refine the practice implementation with the title "individualism in tasks specification"*. The refinement's inputs and outputs are summarized in table (1). For improvements in the refined (i.e. target) practice, the table presents briefly actions (1) and (2). Action (1) is about implementing a practice method for integrating the flow model based analysis and dialogue practices for formative assessments (see section 3 for more details), whereas action (2) is about refining the dialogues, organized for practicing Hattie's pedagogy in visible learning, to include a technical solution (i.e. the specification tool) for visualizing skill mastery goals and assessment feedback for each individual student. The refined practice in table (1) is an example of our dialogue practices for individual support. Fig. 2 highlights the initial design of the dialogue practices in context of five key strategies for formative assessment proposed by Wiliam [17]. As seen in Fig.2, for practicing Hattie's pedagogy the dialogue practices require a tool to produce and individualize reports for two purposes: visualizing a mastery goal, and visualizing assessment feedback during dialogues. As seen in Fig. 2, each key strategy has a brief description and a numeric identifier. The strategy 5 "activating students as learning resources for one another" is not for learning as individual process but for learning as social process. Therefore it is not included in this paper because this paper is a bout refining our practices for individual support. To improve our practices for individual support, strategies 1, 2, 3 and 4 are interpreted and implemented by us via putting four practices in action; namely two dialogue practices and two automated practices. The automated practices are delivered via our specification tool (see section 4), whereas the dialogue practices for reflection are basically guided by Freire's dialogic pedagogy as seen in Fig.3.

TABLE 1. EXAMPLE: REVIEW AND REFINE A PRACTICE IMPLEMENTATION

Practice title: individualism in tasks specification **Goal:** given a previous implementation of the practice, refine it and suggest improvement actions for its new implementation	
Inputs	
Previous practice	• Students have different backgrounds, different habits and cultures, and different prior skills. • **Students' diversity is problem or asset?**
	Implementation Summary: The diversity was seen as asset and not as problem and therefore as response to diversity. No actions were planned for specifying the individual tasks.
Outputs	
Refined practice	**Summary of refinement:** In context of our workshops, students' diversity must be seen as problem and not as asset. The differences and gaps between students are very difficult to be understood by trainers. To understand student diversities, **dialogues with each student should be planned**.
	The required actions for improvements are: **Action (1):** For understanding student engagement and emotion states, a **practice method** for integrating the flow model based analysis and **dialogue practices** for formative assessments should be implemented (more details in Fig. 3). **Action (2)**: **dialogue practices** for Hattie's visible learning pedagogy must include **a technical solution** for visualizing skill mastery goals and assessment feedback for individual (more details in Fig. 2)

Fig.3 summarizes our practice method for integrating the flow model based engagement analysis and analysis of formative assessment outcomes in Freirean dialogues. By enabling Freirean dialogues in ourworkshops, the method enables the student to initialize goals based *on his/her discovery interests and everyday experiences*, and then the method enables the formative assessing of individual progressive challenges for a student based on analyzing his/her engagement (i.e. emotional states) using the flow model (see section 3) .

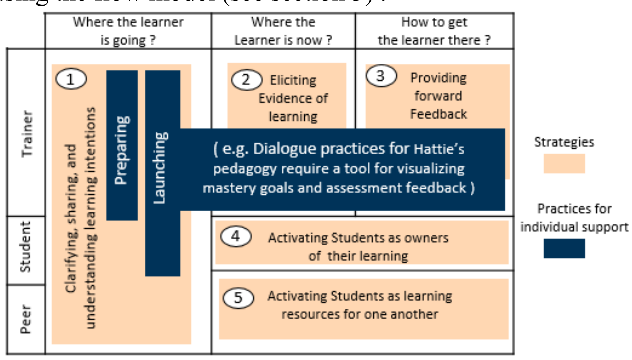

Fig. 2 Initial design of our dialogue practices for individual support in context of the five **key** strategies for formative assessment by Wiliam [17].

As seen in Fig.3, The dialogue is an experimental medium to enable practicing two types of analysis: namely, analysis of formative assessment outcomes (e.g. skills level and skills availability patterns for a student), and analysis of student's emotional states (i.e. engagement) using the flow model. It is useful to summarize why we select Freire's dialogic pedagogy in Fig.3 as initial basis to guide our dialogue practices for reflection. Freire justified the need for dialogue in pedagogy ***epistemologically*** by tracking the origin of reflective knowledge in dialogue and ***ontologically*** by tracking humanity in dialogue. According to Freire, communication of freely participating people, who cannot overrule themselves by authority or power, generates a special type of reflection, which only humans possess. This type of reflection is knowing about ones' own knowledge. This communication (i.e. dialogue) also generates special social relations of humanity among the participants.

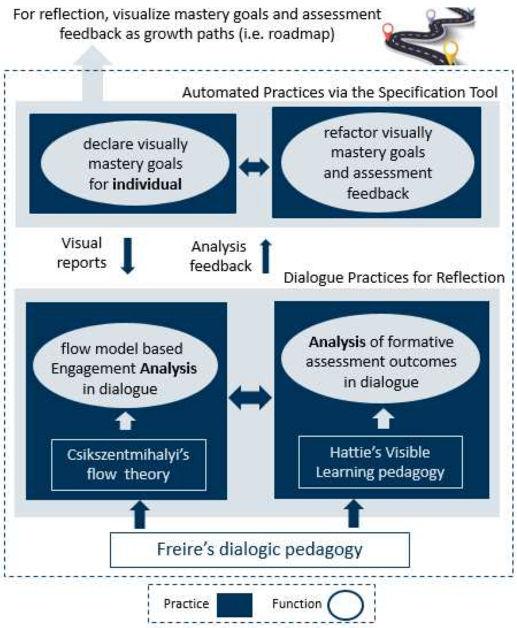

Fig. 3 our practice method for integrating our dialogue practices for reflection, and the automated practices delivered via our specification tool.

Thus, human reflection (episteme) generates humanity, while humanity generates the reflection (i.e. one's own knowledge about his/her knowledge) [18, page 76]. Freire's last justifications (i.e. his epistemological and ontological arguments) hold in our workshops context and we agree with him that dialogue in pedagogy is very important medium to discover the reflective knowledge that distinguishes human beings. To sum up, for discovering reflective knowledge in our workshops, enabling the trainers to produce and individualize visual reports in dialogues is required in context of practicing Hattie's visible learning pedagogy. Hattie's pedagogy informs and guides our practice for dialogue based formative assessments as seen in Fig.3. The practice is about eliciting evidences of learning and clarifying assessment feedback via visual reports, produced by our tool, to support student moves forward in learning and skills mastery. In next section, we present in details the flow model based analysis.

III. Flow Model Based Analysis

As guideline to define the challenges and skills of a goal in a measurable style, we select the flow model. The flow model is explained by Csikszentmihaly's research outcomes as seen in Fig. 4 [13, page 72]. Csikszentmihaly's research is about measuring optimal everyday experiences (e.g. flow experiences) and their negative/positive emotional states in everyday life. The student everyday experiences play an important role in defining the initial goal according to Freire's dialogic pedagogy. Both Freire's dialogic pedagogy and Csikszentmihaly's flow theory are about understanding everyday experiences, therefore an architectural decision was taken by us, in our research based architecting activities, to integrate them in our method as seen in Fig. 3. Fig. 4 summarizes the research outcomes in the map of everyday experience. The map has eight emotional/mental states, namely; control, boredom, apathy, worry, anxiety, arousal, and flow. A concrete example of everyday activity is car driving. When driving a car, people are most often in control state, provided they are not stuck in a traffic jam, in which case they would be in boredom or apathy; or if there is no sudden snowstorm, in which case they are likely to be in anxiety.

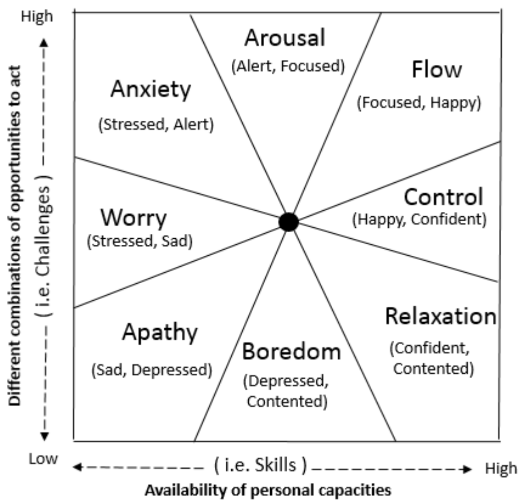

Fig. 4 Flow Model; the Map of Everyday Experience. When People perceive themselves to be above their own personal average level of challenges and skills, they experience flow. The opposite is the state of apathy, where both challenges and skills are low. Other combinations of challenges and skills produce feelings of worry, anxiety, and arousal (when challenges outweigh skills), or control, relaxation and boredom (when skills outweigh challenges). Some of the other prominent emotions typical of each "state" are indicated in parentheses [13, page 72].

The car driving example is used by us as metaphor to explain to the student how we use the flow model in our offers to support him/her in finding individual growth paths (roadmap) for skills mastery (a roadmap is shown in Fig. 3). The dialog practices for flow model based analysis (see Fig.3) are structured by us around the flow model's two dimensions and the central point shown in Fig. 4. The two dimensions and the central point are explained and interpreted in [13] for many everyday experience contexts, but the context of everyday "programing task" experiences is not included, therefore for this context we summarize our interpretation as following.

The central point represents the weekly average level of challenges and skills [14, 19] in programing task experiences. This weekly average level is individual and personal. That means it must be estimated for each student separately because it depends on student personal attributes like everyday habits, background, and prior skills. The estimation is based on scoring criteria in experience sampling method [14, 19]. In brief, it is about estimating the weekly average of student performance score in knowing-how to align (or to balance) challenges and skills based on colleting evidences, during a week, about his/her everyday "learning" or "programing task" experiences. The closer a student's score to this central point, the more average his/her mood tend to be - neither positive no negative. However, as his/her score moves away from this central point, distinct states of mind begin to emerge depending on the ratio of challenges and skills. Basically, the more a person feels skilled, the more her moods will improve; while the more challenges that are present "for immediate goal or task at hand", the more her attention will become focused and concentrated. As we would expect, optimal experience is represented by the flow "state" in Fig. 4 where both challenges and skills lie above the average level; at such moments of optimal experience one is both happy and focused .

The challenges dimension represents "**different combinations of opportunities to act**". We provide the opportunities as outcome of our "programing task" specification. They are provided in many specification forms (i.e. descriptors) that facilitate a very high ceiling of student challenges for skills mastery improvement. In these forms, the student challenges are not only clear and actionable but also easy to be combined and selected by the student in context of activating him/her as owner of his/her learning. For concrete example of these specification forms, see sections 4 and 5. Moreover, the challenges dimension has scale levels between minimum level (i.e. low) and maximum level (i.e. high), where the average level, between the maximum and the minimum, refers to the central point as seen in Fig. 4.

The skills dimension represents "the availability of personal capacities". The availability is about timing and attention resource in one's psychological capital. Psychological capital is a metaphor to simplify explanation: It is useful to think of **enjoyment** as the psychological equivalent of building capital, and of **pleasure** as the equivalent of consumption. As used in economics the term "capital" would be defined as follows: Capital refers to resources withheld from immediate consumption in the expectation of greater future returns. [13, page 76]. To put it differently, "**enjoyable** activity" builds one's psychological capital, whereas "**pleasant** activity" consumes the psychological capital. But what exactly, is the resource at psychological level that will be "consumed" during a "pleasant activity", however in an "enjoyable activity" it will be withdrawn from consumption? The answer to last question is "attention resource". Attention is the brain's capacity to process information and to direct action. It is a limited resource, because we cannot process more than a few bits of information at any single moment, and thus we can only be aware of a tiny fraction of what is going on inside us or around us. Attention is psychic energy, and like physical energy, unless we allocate some part of it to the task at hand, no work gets done. The skills dimension has, like challenges dimension, scale levels with minimum level, maximum level and average level that refers to the central point as seen in Fig. 4. Collecting and evaluating feedback about the availability of a student's personal capacities, to help us specify individual and enjoyable "programing tasks or activities" are conducted in context of dialogue practices (see Fig. 3). To sum up, the model's 8 emotional states in Fig. 4 provide a serviceable compass for finding one's way through the thickets of the emotional life. The direction toward which the model clearly points is flow state. Beside flow state there are two states in Fig. 4 associated with positive emotions. The first is the one labeled "arousal" where due to slightly higher challenges a person must concentrate but does not feel quite at ease. In this situation, if one wants to enter flow, the level of skills must be improved. Because flow state is attractive, a person in arousal is likely to be motivated to reach that state, and thus learns and grows to accomplish that. The other positive state is "control" where skills slightly outweigh challenges. This "control" state does not require high concentration and therefore one does not operate at 100 percent capacity, and so it is not as enjoyable as flow. However, it is relatively easy to move from this state or condition into flow, by somewhat higher challenges. Arousal and control are both states that lead to ***learning*** because they ***stimulate a student to develop higher complexity of skills***. Via our individual support, we aim to create conditions that help student ***experience arousal, control and flow.*** To enable creating these conditions , our API-feature specification tool provides two functions to produce and individualize visual reports. The tool's functions are presented in details.

IV. Tool's Two Functions in Details

Fig. 5 shows how the trainer uses the feedback in a dialogue for initializing and/or refining the input parameters: input 1 for declaring function, and input 2 for refactoring function. Like input1, input2 consists of two parameter sets. Parameter set for a **M**astery **G**oal (i.e. MG parameters) and parameter set for a **F**low **M**odel **S**tate (i.e. FMS parameters).

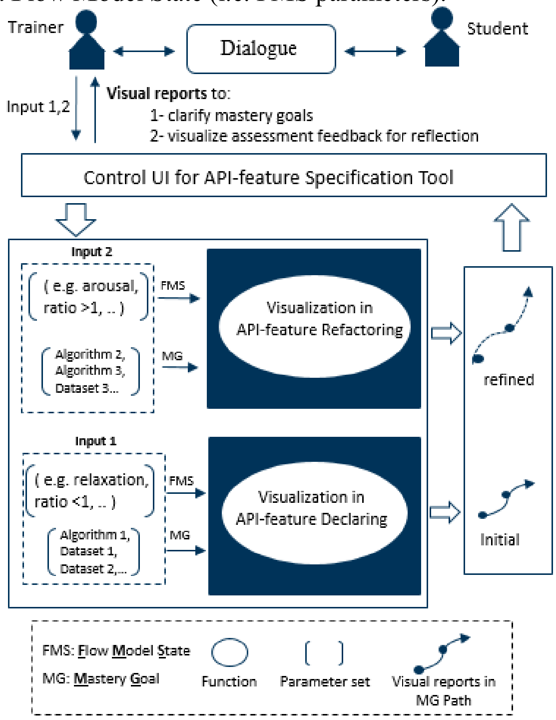

Fig. 5 Tool's two functions in details

Moreover, Fig. 5 shows the main output or aim of the tool. The main aim is to enable trainers produce and individualize visual reports for two purposes namely: to clarify goals for learning and mastering algorithms, and to visualize assessment feedback for reflection and action. The tool is for task specification. By task specification we mean: *a mastery goal specification (i.e. MG parameter set) that has one or many of algorithm specifications and one or many of dataset specifications.* Fig. 5 shows example values of MG parameters and FMS parameters in the declaring function's input. The MG values (algorithm1, dataset 1 and 2) are used to initialize challenges in a mastery goal, whereas the FMS values (relaxation state. ratio of challenges and skills <1 …) are used to initialize a flow model state for a student. Moreover, for the refactoring function's input, Fig. 5 shows also example values. The MG values (algorithm 2, and 3, dataset 3 ...) are used to refine or refactor the challenges in a mastery goal, whereas the FMS values (arousal state, ratio >1 …) are used to refine a flow model state for a student. As Fig 5 shows; the declaration functions' outcome is a visual report in a student's initial learning path, whereas the refactoring function's outcome is a visual report in the student's refined path for learning and mastering algorithms. The FMS parameters are about diagnosing a student's learning engagement, whereas the MG parameters are about the algorithms and datasets used to specify a mastery goal and its progressive challenges. In next section, we present the tool testing and verification.

V. Tool Testing and Verification.

A. Tool Boundary : Metrics for Measuring Effectiveness

The tool boundary is the starting point for determining metrics for measuring effectiveness**.** In context of operating our existing system of tools and concepts (see [7, 8]), the tool is bounded as a self-contained micro system. By self-contained we define operational conditions for enabling two metrics for measuring its effectiveness as it follows: the tool operates using only one text input file **(** input 1 or 2 in Fig.5 shows concrete and compact content of an input text file). The file input format is based on the API standard specification Swagger [20] (more details about swagger specification is in our previous work [7]). To produce a single visual report, given an input text file edited by a trainer, we select two metrics for measuring its effectiveness; namely.

- **Metric for time-on-function use**. It measures time requires to produce a visual report, given an input text file. It has scale from 1 millisecond to 60seconds. For the declaration function, it must be in milliseconds, whereas for the refactoring function it must be in seconds and less than 1 minute because most visual reports in refactoring require a training phase of a machine-learning algorithm (concrete examples in section 6).
- **Metric for lostness**. It measures how lost the trainers are when they use the tool. Lostness scores range from zero to one. A high score (closer to one) means that trainers are very lost and having trouble finding what they need. Because by design the tool operates by using only one declarative input text file, the lostness metric for the tool's both functions must be less than 0.3.

Beside the last two metrics, we have also a non-metric effectiveness measure defined for **maximizing clarity in trainer-student dialogue.** Given a visual report intervention in a dialogue. By definition the intervention goal is to **maximize clarity** in goals communication for understanding and mastering deep-learning principles and algorithms. **(**For more details see test objective and test steps in table 5)**.**

B. Test Setups and Verification Plan

The prerequisites required for setting up the tests are: big datasets and algorithms. In table (2) we present four datasets, used as examples in the workshops. Datasets 1 and 2 are selected to enable students understand the principles and algorithms for shallow-learning, whereas the datasets 3 and 4, to enable them understand the algorithms for deep-learning.

TABLE 2. FOUR BIG DATA SETS USED IN WORKSHOPS

id	Name	Description
1	CO dataset [7]	It is about exam scores for 129 students in Computer Organization (CO) course. The CO dataset was used by us in [7, 8] to predict binary classes (success or fail) or to predict multi classes (Excellent, good, middle, fail). for more info please see [7, 8]
2	Iris dataset [21]	It has 150 instances. Each instance has 4 attributes. It has 3 classes: { setosa, versicolor, virginica}
3	MNIST dataset [22]	MNIST (Modified National Institute of Standards and Technology) dataset. It is about 28x28 grayscale images for handwritten digit from 0 to 9.
4	IMDB dataset [21]	It is a set of 50,000 highly polarized reviews from Internet Movie Database. It is split into 25,000 reviews for training and 25,000 reviews for testing;

Examples of algorithms for shallow-learning deep-leaning are listed in table 3. In setup we have two types of trainer-student dialogues; initial dialogues (for initial specification) and non-initial dialogues (for co-specification after assessment feedback exchange). Table 4 shows samples for freshman students participated in the workshops. They are freshmen students in computer science, software engineering, and medical engineering. The table shows the outcomes of the four initial dialogues with them: 2 of them have **indirect** culture whereas the other two have **direct** culture. The culture based expressing styles "indirect" and "direct" are analyzing terms defined by us for better analyzing and understanding some communicating problems with students: the term "indirect" means the student was educated in previous primary schools to avoid expressing the point or his/her needs, whereas the term "direct" means the student was educated to express directly the point or his/her needs. As described in section 2 (see Fig.3) an initial dialogue has two analysis tasks. As outcomes of the analysis tasks we captured the next variables.

{CA, SA}: Challenges **A**verage, **S**kills **A**verage. The weekly average level of challenges and skills in programing task experiences.

{CL, SL}: **C**hallenges **L**evel, **S**kills **L**evel.

FMS : Estimated current **F**low **M**odel **S**tate.

TABLE 3. ALGORITHMS USED IN WORKSHOPS

algorithms type	Name of algorithms
Sallow-learning	OneR [24] ,k-nearest neighbors [25] , Naive bayes [26] and , C4.5 algorithm [27]
Deep-learning	Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN)

TABLE 4. OUTCOMES OF INITIAL DIALOGUE WITH 4 STUDENTS

id	{CA, SA},{CL, SL}, FMS for CA,CL,SA,SL has scale between 1:min and 9:max	culture
1	{CA< 3.2, SA<1}, FMS= worry	Indirect
2	{CA< 3.2, SA<1}, FMS= worry	Indirect
3	{CA>4.5, SA>6}, FMS= control	Direct
4	{CA<3.2, SA>4.5}, FMS= anxiety	Direct

Given the input variables ({CA, SA}, {CL, SL}, FMS, culture) for a student, the student challenges must be specified in test cases. For two test cases (i.e. for mastery goals A and B), table 5 has a verification plan.

TABLE 5: VERIFICATION PLAN FOR TWO MASTERY GOALS

<table>
<tr><td colspan="5">Mastery goal (A): understanding algorithm performance and principles of shallow-learning and deep-learning.
<u>Student Challenges</u> in understanding
1: the training and testing phases in machine learning
2: the performance comparison between algorithms in shallow-learning
3: shallow-learning vs deep-learning
Mastery goal (B): using deep-learning neural networks to solve classification problems of images and text documents.
<u>Student Challenges</u> in understanding
1: the concepts "layer " andnetwork of layers (i.e. a model)
2: different options for loss functions (images classifier)
3: different options for optimizers (images classifier)</td></tr>
<tr><td colspan="5">Test objective: for each student 1, 2, ... 4 in table 5, use the tool's functions to produce and individualize visual reports <u>to maximize clarity in trainer-student dialogues,</u> given the student different education backgrounds, different culture backgrounds, and different prior skills.</td></tr>
<tr><td></td><td>Id</td><td>Test step details</td><td>Test results</td><td>student id</td></tr>
<tr><td rowspan="5">Mastery goal A</td><td>a.1</td><td>Initialize API-feature specification for mastery goal A, based on initial dialogue outcomes in table 4</td><td>See Fig. 6</td><td>1, 2, 3</td></tr>
<tr><td colspan="2">Analysis (1) in dialogue practice</td><td colspan="2">FMS = "worry" for students 1 and 2. For more analysis's outcomes see section 6</td></tr>
<tr><td>a.2</td><td>Produce report for visualization to reduce the " worry" feelings</td><td>See Fig. 9 and 10</td><td>1, 2</td></tr>
<tr><td colspan="2">analysis (2) in dialogue practice</td><td colspan="2">FMS = "anxiety" for student 2. For more analysis's outcomes see section 6</td></tr>
<tr><td>a.3</td><td>Adjust report for visualization to reduce "anxiety" feelings.</td><td>See Fig. 11</td><td>2</td></tr>
<tr><td></td><td colspan="4">Repeat steps a.1... a.3 based on the changing needs of each student</td></tr>
<tr><td rowspan="3">Mastery goal B</td><td>b.1</td><td>Initialize API-feature specification for mastery goal B, based on initial dialogue outcomes in table 4</td><td>See Fig. 7 and 8</td><td>4</td></tr>
<tr><td colspan="2">analysis (3) in dialogue practice</td><td colspan="2">FMS = "anxiety" for students 4. For more analysis's outcomes see section 6</td></tr>
<tr><td>b.2</td><td>Adjust visual reports to help student reduce "anxiety" feelings</td><td>See Fig. 12</td><td>4</td></tr>
<tr><td></td><td colspan="4">Repeat steps b.1... b.2 based on the changing needs of each student</td></tr>
</table>

To run the tests, table 5 structures briefly the test objectives and test steps. In next section, we present the test results and discussion.

VI. Test Results and Discussion

The test objective and test steps in table (5) are driven by the non-metric effectiveness measure: maximizing clarity in trainer-student dialogues as individual support for each student. Given a visual report intervention in a dialogue, by definition the trainer's intervention aims to maximize clarity in goals communication for understanding and mastering deep-learning principles and algorithms. indvidual support In for each student we classify the dialogues as initial dialuges to enableinitial specifications for clarifying mastery goals, and non-initial dialogues to enable co-specifications to visualize assessment feedback for reflection. Each dialogue might include at leastone of the two analysis tasks see) Fig.3 section 2 and). Given the analysis feedback and outcomes in a dialogue as summarized in tables 4 and 5, we present the test (i.e. intervention) results in initial and co-specifications as it follows.

A. Initial Specifications to Clarify Mastery Goals.

Fig. 6 shows the test result for step a.1, whereas Fig. 7 and Fig. 8 show the test results for step b.1 in table 5. Each of Figs. 6, 7 and 8 is produced by the declaration function in less than 1 second (i.e. time-on-function use is about 1 second).

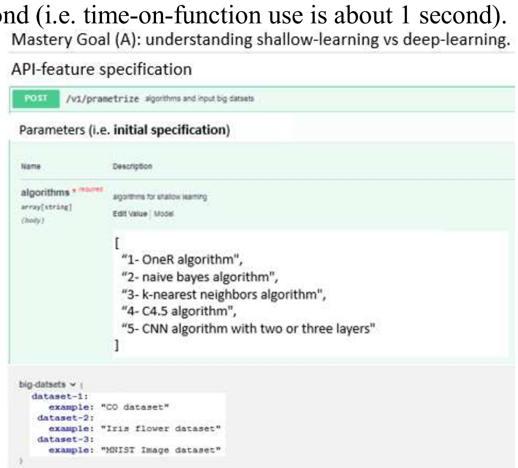

Fig. 6 test results for step a.1 in table 5

For students 1, 2 and 3; given the initial analysis outcomes in table 4, students 1, and 2 are in "worry" state and they have beginner programming skills, whereas student 3 is in "control" state and he has intermediate skills. As responses to the available skills for each student, and to his/her initial emotional state, the trainer produces in Fig. 6 a visual report to clarify an initial specification for the mastery goal (A). The goal (A) is defined to practice how algorithms for shallow-learning and deep-learning work and what are their differences. Briefly, Fig. 6 visualizes specifications for five algorithms and three datasets. As flexible options to initialize the algorithms, the datasets, and a combination of them to fit a student's changing need. Fig. 6 shows visually the flexible options for four shallow-learning algorithms namely: OneR, naive bayes, k-nearest neighbors, and C4.5 and one CNN algorithm for deep-learning. It shows also flexible options for three input big datasets namely; CO dataset, iris flower dataset and Mnist image dataset. For students 4, an initial specification for the mastery goal (B) is presented in Fig. 7 to fit his initial needs. He has advanced programming skills and therefore the goal (B) is about understanding how to develop deep-learning networks to solve the problems of images classification. The report in Fig. 7 has a supportive diagram (see Fig. 8) to clarify the terms: layers, networks of layers, loss functions and optimizers. Fig. 8 clarifies the anatomy of a neural network: to train a neural network the following objects are needed: ***Layers***, which are combined into *a network (or model);* the *input data* (X) and corresponding *targets(Y); the loss function* defines feedback signal for machine learning, and the optimizer determines how it proceeds.

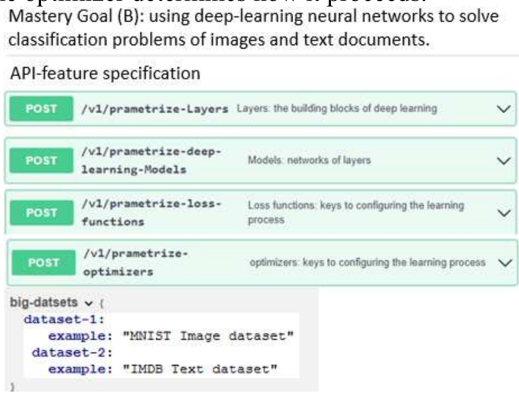

Fig. 7 Test result for step b.1 in table 5

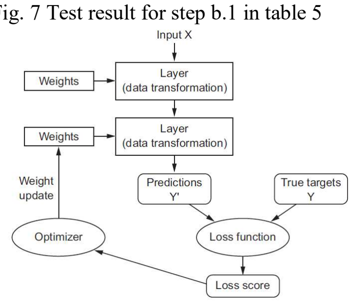

Fig. 8 supportive diagram for report in Fig. 7. Adapted based on [28]

Fig. 8 shows the interaction between these objects: the network, composed of layers that are chained together, maps the input data (X) to predictions (**Y'**). The loss function then compares these predictions (**Y'**) to the targets (Y), producing a loss value: a measure of how well the network's predictions match what was expected. The optimizer uses this loss value to update the network's weights (i.e. its trainable parameters).

B. Co-specifications to Visualize Assessment Feedback.

As next step in individual support (i.e. next intervention) for student 1, Fig. 9 shows a report produced after a dialogue based assessment with him. The assessment feedback is visualized to explain to him why he is in a "worry" state.

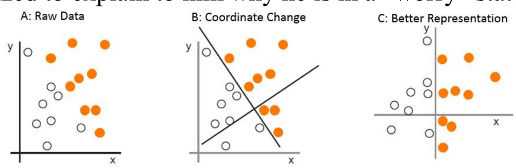

Fig. 9. Test result for step a.2 in table 5

The main reason was misunderstanding the meaning of the term "learning" in machine learning. Report in Fig.9 highlights that the central problem in machine-learning is to ***meaningfully transform data***: in other words, to learn useful representations of the input data at hand. Fig. 9b and Fig. 9c show simple and concrete data transformations to get useful and better representation, with this representation, the orange/white classification problem can be expressed as a simple rule: "orange points are such that $x > 0$," or "White points are such that $x < 0$." This new representation basically solves the orange/white classification problem. Student 2 has "**worry**" feelings because of different reason. The reason is; for her how to use the metrics for **algorithm** accuracy and detailed accuracy were not clear. To reduce her worry a report in Fig. 10 is produced and individualized by the refactoring function. The report is a **positive intervention** to support her in understanding how to use the metrics for a classifier algorithm accuracy and for detailed accuracy (like precision, recall and F-measure) during training and testing phases of a machine-learning algorithm (like C4.5 algorithm). The C4.5 algorithm and the CO dataset in Fig.10 were very effective to provide suitable and progressive challenges to her. The support aims to help her understand the simple prediction and classification algorithms. Student 2's emotional state changing and her skills availability are captured by our flow model based **engagement** analysis and the analysis of our dialogue based formative assessments during the individual support process. Via the analysis feedback we elicit evidences that her mood was moved from "**worry**" to "**anxiety**" because she cannot understand how to attack a prediction problem in deep-learning based image classifier. To help her reduce the "**anxiety**" report in Fig. 11 is individualized for her.

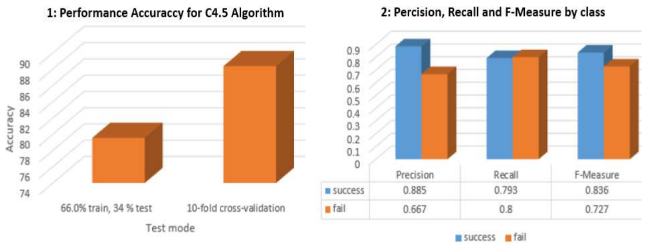

Fig. 10: Test result for step a.2 in table 5

The report in F.11 was also a positive intervention to start feedback exchange and reflections with her. With her we share the reflections and the actionable feedback as a set of systematic explaining and action steps. The steps are:

Step1: The main reason for her anxiety is explained and visualized in Fig. 11b [28]. Fig. 11b shows the complexity of digit-classification model required to solve the prediction problem in deep-learning based image classifier selected by student 2. She selected over challenging task in comparing to her skills and that was the main reason for her anxiety.

Step2: In Fig. 11a we show her a decision tree based model using the CO dataset, to predict if a student can pass or fail the final CO course exam. Because Co course was included in her study plan, developing and understanding prediction models based on the CO dataset were very suitable and moderately increasing challenges for her.

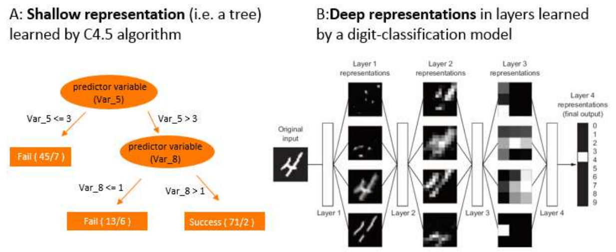

Fig. 11: Test result for step a.3; visual report individualized for student 2

Step3:Last feedback for her was to develop prediction models using the simple Iris image dataset presented in table 2 by applying the decision tree based algorithms, like C4.5, she applied in the CO dataset. Like CO dataset, the Iris dataset is designed for shallow-learning algorithm. The last three steps were very effective as actionable feedback to help her change her mood from "**anxiety**" to "**control**" state.

Visual reports in Figs 10 and 11 are positive interventions to student 2, however for student 4 the visual reports for algorithm accuracy produced during the individual support process with him were ***negative*** interventions. A ***positive*** intervention for student 4 is presented in Fig.12. In Fig.12 a python snippet is shared with him to discuss his needs and to help him move from "**anxiety**" to "**control**" state. The main reason for the positive impact of sharing this python snippet is because he has good programing skills in python language. In this python snippet, a model for full-connected neural network is presented. The model has two simple "Dense" layers. For student 4, comparing the model's accuracy using the loss functions: MAE [29] and MSE [30] has positive impact not via producing visual reports but via pair-programming sessions with him. For him communicating goals ***verbally*** in co-programming are more effective than communicating goals in co-specification like for student 2.

```
#------- load the MNIST dataset
(train_images, train_labels), (test_images, test_labels) = mnist.load_data()
#------- Converts a class vector (integers) to binary class matrix
num_classes = 10
train_labels = keras.utils.np_utils.to_categorical(train_labels,num_classes)
test_labels = keras.utils.np_utils.to_categorical(test_labels,num_classes)
#------- Create a Model for Fully Connected Neural Network
network = Sequential()
network.add(Dense(units=512, activation='relu', input_shape=(28 * 28,)))
network.add(Dense(units=10, activation='softmax'))
network.compile(loss='MAE', optimizer='rmsprop',
                metrics=['accuracy'])
#--------Train the Model
network.fit(x=train_images, y=train_labels, batch_size=128, epochs=5)
```

Fig. 12: Test result for step b.2 in table 5;

By our definition to understand student background culture (see table 4 in section 5), student 2 belongs to the ***indirect expressing*** culture, whereas student 4 belongs to ***direct expressing*** culture. Culturally student 4 is a product of direct culture therefore it is easy for him to express directly via ***verbal actions*** his needs and goals, whereas for student 2 direct expressing of her needs and goals is a skill that she has to lean beside learning and mastering algorithms skills. For her this increases the cognitive load and reduces her engagement to understand and master the algorithms. If visual reports interventions are not shared with her, her disengagement with high probability is likely to happen. Because her emotional state cannot be moved from "anxiety or worry" to "control". As seen in table (6), for students in indirect2 entculture group (like stud) the visual reports interventions (like Figs 6...11) maximize clarity in goals communication for understanding and mastering the machine-learning based classifier algorithms, whereas for students in directculture group (like studnet 4) visual reports interventions have minor effectiveness or have very low probability in maximizing the clarity in goals communications. What works for this group is ***verbal*** interventions without visual reports produced by the tool. ***Given two culture groups (direct and indirect)*** of students, table 6 summarizes the test outcomes for the non-metric effectiveness measure: maximizing clarity in trainer-student dialogues. For the metric effectiveness measures "time-on-function use" and "lostness" table 7 summarizes the results.

TABLE 6. MAXIMIZING CLARITY IN TRAINER-STUDENT DIALOGUES

Culture group	Interventions **visually via the Tool**	Intervention **verbally without Tool**
Indirect (8 students)	Very effective	Not effective
Direct (6 students)	Minor effectiveness	Very effective

The results in table 7 show that the tool's two functions are effective as expected by our design objectives (see section 5.a)

TABLE 7. RESULTS OF METRICS FOR MEASURING EFFECTIVENESS

Metric	Declaration function	Refactoring function
time-on-function use (measured for 30 different input text file)	Max measured value is 9 milliseconds. **It is less than 10**. 10 by design is the max allowed value.	Max measured value is 25 seconds; **it is less than 60**. 60 by design is the max allowed value.
Lostness (measured for 4 trainers)	Max measured value is 0.1. **It is less than "0.3"**. "0.3" by design is the max allowed value	

The student feedback if the shared visual reports contribute positively or negatively to maximizing the clarity in goals communication is also collected. 15 of 16 students agree that sharing the visual reports is helpful, whereas only one student disagrees. To sum up, for students in indirect culture (like students 1, and 2), the test results in Figs 6..12, and in tables 6 and 7, show the effective use of the tool to create conditions to overcome the student's negative emotions (e.g. worry and anxiety), and help him/her experience positive emotions (e.g. control) in learning and mastering the deep-learning based classifier algorithms. To put it differently, the results show the effective use of the tool to create situations in initial specification and co-specification of tasks, where student's skills are engaged by progressively higher challenges to understand the deep-learning based classifier algorithms. However, when it comes to the direct culture group of students (like student 4), the tool has minor effectiveness. For this group of students, the verbal interventions in two-person dialogue (i.e. trainer-student dialogue) is effective provided that trainer time resources are available. Not only for two-persons dialogue settings, but also for settings that include more than one student in dialogue, we aim to further design our tool in future. Future works and conclusions are presented in next section.

VII. CONCLUSIONS and Future Work

The innovative design of a trainer's support tool for specifying deep-learning API features is presented. The design aims to support the trainers in practicing Hattie's Visible Learning [1] pedagogy in dialogues for tasks specification and formative assessments during workshops. The design enables two support functions for visualization in declaring and refactoring the API-features. The functions are put under tests and the test results show that they can be effectively used by the trainer to produce and individualize visual reports for two purposes:

1. To clarify goals for learning and mastering algorithms. The reports clarify goals with progressively higher challenges to understand deep-learning principles and algorithms.
2. To visualize assessment feedback in dialogue practices. The assessment feedback visualized by the reports enables reflections for action. This maximizes clarity in goals communication to better understand the deep-learning algorithms.

Moreover, to understand student engagement for better specifications, our practice method is presented. The method is for integrating the flow model based engagement analysis, Hattie's pedagogy and Freirean dialogue based formative assessments. Not only for two-persons dialogue settings, but also for settings that include more than one student in dialogue, we aim to further design and improve our tool in our future research work as software architects. Because of the limits in trainer resources who have expertise in putting Hattie's pedagogy and the flow theory in action, scaling out the tool and the flow model based analysis for big number of students is very difficult and challenging. Solving the difficulty of scaling out the tool is planned in our future work. Moreover in future work, the systematic steps of the flow model based analysis will be restructured to enable automatic detecting of student emotional states using techniques in text based emotion recognition. Last but not last, the theoretical views for ***personal best goal*** as defined by Martin, A.J et al in [31] is also planned to beinvestigated .in our future work

REFERENCES

[1] Visible Learning for Teachers: Maximizing impact on learning, John Hattie, First published 2012 by Routledge, 2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN, and Simultaneously published in the USA and Canada, by Routledge, 711 Third Avenue, New York, NY 10017.

[2] NACCCE: All Our Futures: Creativity, Culture and Education (1999).

[3] Hattie on Assessment-capable learners, https://www.youtube.com/watch?v=fen8M1_MfvE, retrieved on 13.06.2022.

[4] Robert E. Stake (born December 18, 1927) is a Professor Emeritus of Education at the University of Illinois, Urbana-Champaign. https://en.wikipedia.org/wiki/Robert_E._Stake, or via Utube https://www.youtube.com/watch?v=_bf59TDochY , retrieved on 10.04.2022.

[5] Hattie J, Timperley H. The Power of Feedback. Review of Educational Research. 2007;77(1):81-112. doi:10.3102/003465430298487

[6] The Turning Point for the Teaching Profession; Growing Expertise and Evaluative Thinking, By Field Rickards, John Hattie, Catherine Reid, ISBN 9780367531867, Published November 30, 2020 by Routledge.

[7] W. Mahfouz and H. -D. Wuttke, "Big Data Analytics APIs Architecture for Formative Assessors," 2021 IEEE Frontiers in Education Conference (FIE), 2021, pp. 1-9, doi: 10.1109/FIE49875.2021.9637431.

[8] W. Mahfouz and H. Wuttke, "Automatic Classifiers for Formative Assessment," 2019 IEEE Frontiers in Education Conference (FIE), 2019, pp. 1-9, doi: 10.1109/FIE43999.2019.9028713.

[9] Freire, Paulo (1972), Pedagogy of the Oppressed. Harmondsworth: Penguin, Or via link https://en.wikipedia.org/wiki/Pedagogy_of_the_Oppressed

[10] https://en.wikipedia.org/wiki/John_Dewey , retrieved on 13.06.2022

[11] https://en.wikipedia.org/wiki/Paulo_Freire , retrieved on 13.06.2022

[12] Noam Chomsky On The Perils Of Market-Driven Education https://rozenbergquarterly.com/noam-chomsky-on-the-perils-of-market-driven-education/ , retrieved on 13.06.2022

[13] Mihaly Csikszentmihalyi : Good Business: Leadership, Flow, and the Making of Meaning PENGUING BOOKks, 2003.

[14] Moneta, G.B. (2021). On the Conceptualization and Measurement of Flow. In: Peifer, C., Engeser, S. (eds) Advances in Flow Research. Springer, Cham. https://doi.org/10.1007/978-3-030-53468-4_2

[15] Buckingham Shum, Simon and Deakin Crick, Ruth (2012). Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics. In: 2nd International Conference on Learning Analytics & Knowledge, 29 Apr - 02 May 2012, Vancouver, British Columbia, Canada.

[16] Hattie, John A. C.. "Models of Self-Concept that are Neither Top-Down or Bottom-Up : The Rope Model of Self-Concept" (2004).

[17] Wiliam, D. (2011). Embedded Formative Assessment, Bloomington: Solution Tree Press, (Revised edition 2018)

[18] Matusov, Eugene: Journey into dialogic pedagogy, 2009, ISBN 978-1-60876-969-8 (E-Book), Published by Nova Science Publishers, Inc.

[19] https://en.wikipedia.org/wiki/Experience_sampling_method , retrieved on 5.04.2022

[20] https://swagger.io/, retrieved on 5.04.2022

[21] https://archive.ics.uci.edu/ml/datasets/iris, last retrieved 10.05.2021

[22] http://yann.lecun.com/exdb/mnist/, last retrieved 10.05.2021

[23] https://datasets.imdbws.com/, last retrieved 10.05.2021

[24] https://www.dbs.ifi.lmu.de/~zimek/diplomathesis/implementations/EHNDs/doc/weka/classifiers/rules/OneR.html, last retrieved 10.05.2021

[25] https://www.cs.tufts.edu/~ablumer/weka/doc/weka.classifiers.NaiveBayes.html , last retrieved 10.05.2021

[26] https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm, last retrieved 10.05.2021

[27] https://en.wikipedia.org/wiki/C4.5_algorithm, last retrieved 10.05.2021

[28] https://www.manning.com/books/deep-learning-with-python-secondedition , retrieved 13.04.2022.

[29] https://www.tensorflow.org/api_docs/python/tf/keras/losses/MeanAbsoluteError, retrieved 13.04.2022.

[30] https://www.tensorflow.org/api_docs/python/tf/keras/metrics/mean_squared_error, retrieved 13.04.2022.

[31] Martin, A.J., & Elliot, A.J. (2016). The role of personal best (PB) goal setting in students' academic achievement gains. Learning and Individual Differences, 45, 222-227. DOI: 10.1016/j.lindif.2015.12.014.