

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

## Automatic Formative Assessment Strategies for the Adaptive Teaching of Mathematics

### **This is the author's manuscript**

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/1852671> since 2023-09-19T08:17:12Z

*Publisher:*

Springer Science and Business Media Deutschland GmbH

*Published version:*

DOI:10.1007/978-3-030-86439-2\_18

*Terms of use:*

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

# Automatic formative assessment strategies for the adaptive teaching of Mathematics

Alice Barana<sup>1</sup>[0000-0001-9947-5580], Cecilia Fissore<sup>2</sup>[0000-0003-1007-5404] and Marina Marchisio<sup>1</sup>[0000-0001-8398-265X]

<sup>1</sup> Department of Molecular Biotechnology and Health Sciences, University of Turin, Via Nizza 52, 10126, Turin, Italy

<sup>2</sup> Department of Foreign Languages, Literatures and Modern Culture, Via Giuseppe Verdi, 10, 10124 Turin, Italy  
{alice.barana, cecilia.fissore, marina.marchisio}@unito.it

**Abstract.** Adaptive teaching is defined as applying different teaching strategies to different groups of students so that the natural diversity prevalent in the classroom does not prevent each student from achieving success. Some of the strategies for adaptive teaching are formative assessment, feedback, and self-regulated learning. For the application of these strategies, technology plays a fundamental role. Our research group has developed and tested a model for automatic formative assessment and interactive feedback for STEM using an Automatic Assessment System to allow adaptive teaching. This article presents the design, the implementation, and the analysis of adaptive teaching activities carried out by secondary school teachers of STEM disciplines during a training course. Teachers designed the transformation of questions created for standardized assessment into formative assessment questions, reflecting on how to adapt the requests to students' different needs and how to create guided learning paths. In this paper, all the steps of the training activity carried out will be analyzed using some representative examples: the design of the questions with an analysis of the adaptive teaching strategies adopted; the reflections of the teachers at the end of the course; carrying out activities with students and analyzing their responses; the reflections of the teachers at the end of the activity.

**Keywords:** Automatic Formative Assessment, Automatic Assessment System, Adaptive Teaching, Interactive Feedback, Self-regulation, Teacher Training, Mathematics.

## 1 Introduction

In recent years, educational technology is continually evolving and growing and can be used to integrate school learning experiences and offer adaptive teaching. Thanks to the new technological tools, teachers can offer all students personalized teaching to their different needs and their different cognitive styles [1]. Technology can be an additional resource in the classroom and at home, supporting and helping students in the cognitive, educational, and training process. Indeed, technology can improve learning only if it

helps effective teaching strategies, for example, when it allows to increase the time dedicated to learning, when it facilitates cognitive processes, when it supports collaboration between students, when it allows to apply different teaching strategies to different groups of students or when it helps overcome specific learning difficulties [2].

One of the teaching practices in which technology can offer fundamental support to teachers but also to students is assessment. For a standardized or summative assessment, where the goal is to measure students' learning outcomes, an Automatic Assessment System (AAS) offers the possibility to automatically evaluate, collect and analyze student responses, saving time in correcting tests. The best-known example of national standardized assessment in Italy is INVALSI tests (<https://invalsi-areaprove.cineca.it/>) in the Mathematics, English, and Italian disciplines [3, 4]. From the past three years onwards, the tests for grades 8, 10, and 13 are computer-based, while grades 2 and 5 are still taking place on paper. An AAS can offer fundamental support also for formative assessment. Formative assessment is a process in which students are active protagonists and have the opportunity to understand what has been or has not been learned and how to learn it. Students can also understand the progress made and the difficulties they have in learning [5]. For example, through formative assessment, they can be trained in self-assessment to better prepare themselves for the INVALSI tests by receiving feedback on their work and advice on how to improve.

Our university has successfully developed and tested a model for designing automatic formative assessment activities by using an AAS for STEM and other disciplines [6]. This article presents a training course for teachers focused on the adaptation of questions designed for standardized assessment to questions for formative assessment with the use of an AAS to develop mathematical skills and problem solving in the frame of INVALSI tests. The 8-hour course was aimed at teachers of lower and upper secondary schools within the national Problem Posing and Solving Project, and it was carried out entirely online. 17 teachers attended the course. After 4 weekly synchronous online meetings lasting one hour, teachers were asked to create at least 3 questions with automatic assessment for a formative test (one for assessing knowledge, one for solving a problem, and one for justification) and experiment them with one of their classes. In creating the questions, the teachers reflected on how to create questions that would adapt to the different needs of students and guide them in preparing the tests. The teachers were also asked to answer two questionnaires, one at the end of the questions preparation and one after the experimentation in the classroom. In this paper, the teachers' activities are analyzed and discussed, and representative examples are shown. Teachers' reflections and observations on the proposed training module and the activity carried out in the classroom with the students are also analyzed. This paper is the extension of a paper presented at the 12<sup>th</sup> International Conference on Computer Supported Education [7]. We deepened the analysis of the teachers' questions and examined the students' answers to these questions. The activities carried out with the students and their answers to the teachers' questions were analyzed in detail to study the effectiveness of automatic formative assessment and interactive feedback. The analyses of the teachers' responses to the initial and final questionnaires were also deepened to understand better any difficulties encountered in carrying out the students' activities and their post-experimentation reflections.

## 2 Theoretical framework

### 2.1 Adaptive teaching, formative assessment, and feedback

Every teacher in their class deals with a great variety of students. For example, there may be multicultural classes; students may have different learning styles, individual attitudes and inclinations, or learning disabilities. However, the learning objectives are common to all students. To ensure that all students achieve the same objectives, tailor-made teaching of each student's characteristics, needs, and curiosities can be adopted. Adaptive teaching is defined as "applying different teaching strategies to different groups of students so that the natural diversity prevalent in the classroom does not prevent each student from achieving success" [8]. Some of the strategies for adaptive teaching are formative assessment, feedback provision, and self-regulated learning.

The definition of formative assessment that we adopt is that of Black and Wiliam [9], well known in the literature: "Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded than the decisions they would have taken in the absence of the evidence that was elicited." The authors conceptualize formative assessment through the following five key strategies: clarifying and sharing learning intentions and criteria for success; engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding; providing feedback that moves learners forward; activating students as instructional resources; activating students as the owners of their own learning.

Sadler conceptualizes formative assessment as how learners use information from judgments about their work to improve their competence [10]. Formative assessment is opposed to summative assessment (assessment where the focus is on the outcome of a program) because the former is an ongoing process that should motivate students to advance in the learning process and provide feedback to move forward. In Mathematics Education, summative assessment design is generally affected by psychometric tradition, that requires that test items satisfy the following principles [11]:

1. unidimensionality: each item should be strictly linked to one trait to be measured;
2. local independence: the response of an item should be independent from the answer to any other item;
3. item characteristic curve: low ability students should have low probability to answer correctly to an item;
4. non-ambiguity: the question should be written so that students are led into the only correct answer.

Questions built according to this model are generally limited in the Mathematics that they can assess. The possible problems are reduced to those with one only solution, deducible through the question text's data. If problems admit multiple solving strategies, the only information detected is the students' solution, thus removing the focus from the process, which is essential for assessing Mathematics understanding [12]. Having no information on the reasoning carried out by students, on the resolute

strategies adopted by them and on the registers of representation used, if students provide a wrong answer, it is not possible to understand the type of error made by them and to provide correct feedback. In adaptive teaching, feedback plays an essential role in reducing the discrepancy between current and desired understanding [13]. Effective feedback must answer three main questions: “Where am I going?”, “How am I going?”, “Where to next?”. In other words, it should indicate what the learning goals are, what progress is being made toward the goal, and what activities need to be undertaken to make better progress. Feedback can work at four levels: task level (giving information about how well the task has been accomplished); process level (showing the principal process needed to perform the task); self-regulation level (activating metacognitive process); self-level (adding personal assessments and affects about the learner).

Self-regulated learning is a cyclical process, wherein the students plan for a task, monitor their performance, and reflect on the outcome. According to Pintrich and Zusho [14], “self-regulated learning is an active constructive process whereby learners set goals for their learning and monitor, regulate, and control their cognition, motivation, and behavior, guided and constrained by their goals and the contextual features of the environment.” Self-regulated students become masters of their learning processes.

Technologies can help offer increasingly adaptive teaching: to activate effective formative assessment strategies, give feedback that differs according to level and to students’ responses, and activate self-regulated learning [15].

An Adaptive Educational System uses data about students, learning processes, and learning products to provide an efficient, effective, and customized learning experience for students. The system achieves this by dynamically adapting instruction, learning content, and activities to suit students’ abilities or preferences [16].

Our automatic formative assessment model allows to offer adaptive teaching [17, 18] to assign different activities to students according to their level and promote engagement in Mathematics at the school level [19, 20]. The AAS also allows you to collect and analyze all students’ answers automatically, the results of the checks, and the data on their execution (such as start, end, duration, number of attempts). Thus, the model allows to provide students with high-quality information about their learning and to provide teachers with information that can be used to tailor teaching.

## **2.2 Our model of automatic formative assessment and interactive feedback**

In a Virtual Learning Environment (VLE), formative assessment can be easily automatized to provide students immediate and personalized feedback. Using Moebius AAS (<https://www.digitaled.com/products/assessment/>), our research group has designed a model for designing automatic formative assessment activities for STEM, based on the following principles [6]:

- availability of the assignments to students who can work at their own pace;
- algorithm-based questions and answers, so that at every attempt, students are expected to repeat solving processes on different values;
- open-ended answers, going beyond the multiple-choice modality;

- immediate feedback, returned to students at a moment that is useful to identify and correct mistakes;
- contextualization of problems in the real world, to make tasks relevant to students;
- interactive feedback, which appears when students give the wrong answer to a problem. It has the form of a step-by step guided resolution that interactively shows a possible process for solving the task.

This model relies on other models of online assessment and feedback developed in literature, such as Nicol and Macfarlane-Dick's principles for the development of self-regulated learning [21] and Hattie's model of feedback to enhance learning [13]. The model was initially developed to improve the learning of STEM disciplines, but in recent years, we have also experimented the model for language learning [22, 23].

The use of an AAS also simplifies the analysis of the activities carried out by the students. In Moebius, all the data and results about the students' tests are automatically stored and archived in the AAS gradebook, even in the case of several attempts, and teachers can view them at any time. Each test carried out by the student is characterized by numerous metadata, such as the date on which the test was carried out, the start and end time, the duration, the score of each question and the final one, and the student's data. For each question, statistics are automatically created, including success rate, p-value, number of times students have answered the question (in total, correctly, partially correctly and incorrectly). These data facilitate the teacher's analysis in studying the significance of a question and students' behavior in responding to it.

### 2.3 Adaptive teaching in the national PP&S Project

The model developed for automatic formative assessment is also used within the Italian PP&S Problem Posing and Solving Project of the Ministry of Education [24, 25]. The PP&S Project promotes the training of Italian teachers of secondary schools on innovative teaching methods and creating a culture of Problem Posing and Solving with ICT use. The PP&S Project adopts the following technologies as essential tools for professional growth and the improvement of teaching and learning: a Digital Learning Environment (DLE), based on a Moodle-learning platform, available at [www.progettopps.it](http://www.progettopps.it), integrated with an Advanced Computing Environment (ACE), an AAS, and a web-conference system. The tools used within the PP&S Project support adaptive teaching:

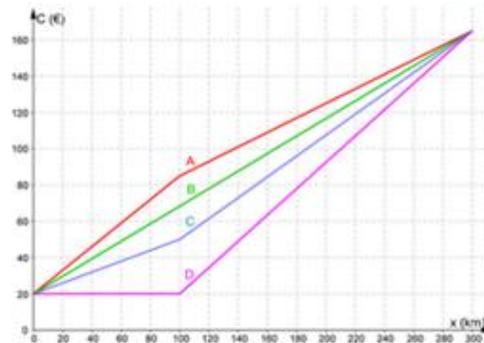
- the DLE allows synchronous and asynchronous discussions, collaborative learning, interactivity and interaction, integration with tools for computing and assesment, and activity tracking;
- the ACE allows interactive exploration of possible solutions to a problem, multiple representations, feedback from automatic calculations, and interactive explorations;
- the AAS allows students to independently carry out the necessary exercises, have step-by-step guided solutions to learn a method, and make repeated attempts of the same exercise with different parameters and values. The AAS promotes students' autonomy and awareness of their skills and facilitates teachers' class management.

The PP&S Project offers various training activities that allow teachers to reconsider their teaching using adaptive teaching with technologies [26]: face-to-face training, online training modules, weekly online tutoring, online asynchronous collaboration, and collaborative learning within a learning community. In this case, we proposed an 8-hour online training module entitled “Automatic formative assessment for preparation for INVALSI tests” to reason and reflect on the adaptation of questions designed for standardized assessment to questions for formative assessment and to create activities with automatic assessment for developing skills in the framework of INVALSI tests. INVALSI test materials were proposed in the training module, created to measure students’ skills, in terms of formative assessment, with the aim of training skills. The INVALSI question repository is full of very valid and interesting questions, which can be used to make lessons in the classroom every day for teaching [27].

### 3 From standardized assessment to formative assessment

Standardized assessment and formative assessment have very different characteristics. In the standardized assessment, each item responds to a single goal; the answer to each item is independent of the answers to the other items; and each question has only one possible correct answer. In formative assessment, problems can have different solutions, answers to items can be dependent on each other, and the solving process should be considered more important than the answer to the question [28]. The following lines present an example of how to transform a question for standardized assessment into a question with formative assessment through the AAS.

We chose an INVALSI question related to the area “relations and functions” in the dimension of “knowing”, conceived for grade 13 students, inserted in a sample test on the INVALSI website. The purpose of the question is to identify the graph of a piecewise function verbally described in a real context. The text of the question is: “A city offers a daily car rental service that provides a fixed cost of 20 euros, a cost of 0.65 euros per km for the first 100 km and a cost of 0.4 euros per km over the first 100 km”. Given a figure (Fig. 1) with the four car rental contracts’ graphs, the student must select the graph corresponding to the proposed offer in a multiple-choice task.



**Fig. 1.** Graphs of four possible car rental contracts [7]. Source: [www.invalsi.it](http://www.invalsi.it).

For the formative assessment, this question can be expanded into three sub-questions of increasing difficulty, keeping the same text of the problem but setting algorithmic values so that numeric values randomly change at every attempt. In this way, students can practice answering several times, always having new data. The first part asks students how much they would spend on a 10 km journey. Students can enter the numerical value (Fig. 2) and have two attempts to see if the answer is correct. The second part asks students to choose the correct graph among four proposed graphs (as in the previous question in Fig.1). The advantage is that each time the problem data change, the four graphs change accordingly. In the last part of the question, students must enter the expression of the function that expresses how the taxi fare varies in the time  $t$  expressed in minutes (Fig. 3). The question proposes three sub-questions of increasing level with three different types of representation of the same mathematical concept (graphic, numeric, and symbolic). Students can test themselves by having immediate feedback. Immediate feedback, shown while the student is focused on the activity, facilitates self-assessment, and helps students stay focused on the task. Moreover, the interactive feedback with multiple attempts encourages students to test themselves and immediately rethink the reasoning and correct themselves.

The screenshot shows a digital assessment interface. At the top, it says "Risposta: Number euro". Below this is a text input field containing the word "Number". Underneath the input field is a blue button labeled "Verify". Above the "Verify" button, the text "Clicca su Verify per controllare la tua risposta e proseguire." is displayed. At the bottom of the interface, it says "Attempt 1 of 2".

**Fig. 2.** Response area to the first sub-question with the “Verify” button [7].

The screenshot shows a digital assessment interface for a symbolic question. It features two input fields, each preceded by a curly brace. The first input field is associated with the condition  $t \leq 10$  and the second with  $t > 10$ . Each input field has a small icon to its right. Below the input fields is a blue button labeled "Verify". Above the "Verify" button, the text "Clicca su Verify per controllare la tua risposta e proseguire." is displayed. At the bottom of the interface, it says "Attempt 1 of 3".

**Fig. 3.** Response area to the last sub-question with the “Verify” button [7].

## 4 Methodology

The training module proposed teachers to analyze the INVALSI math tests' characteristics to create automatic formative assessment activities for developing mathematical skills. The three dimensions of mathematical competence evaluated by the INVALSI tests were analyzed (knowing, solving problems, and justifying) by creating examples of questions for grades 8, 10, and 13. The course did not require prerequisites and was open to all teachers of the PP&S Community, to those who already had experience with the AAS and to those who had never used it. The course duration was 4 weeks and included 4 one-hour synchronous online meetings, carried out through a web conference service integrated with the platform of the Project. In the following months, the teachers were asked to create 3 questions with automatic assessment: one designed to

assess knowledge, one for solving a problem, and one for justification. In particular, the teachers could choose a question with standardized assessment and modify it in a question with formative assessment, reflecting how the transformation was carried out. After that, the teachers experimented the questions with students in one of their classes and shared the questions with the PP&S Teacher Community. The video recording of the online meetings was also made available to all the teachers of the Community.

To understand the training module's appreciation and see how the teachers dealt with the process from standardized to formative assessment, we analyzed the questions created by the teachers and the answers to the two questionnaires. The teachers were asked to answer the initial questionnaire at the end of the four synchronous online training meetings. In this questionnaire, the teachers were asked if they had already used the AAS, if they appreciated different aspects of the training course and the proposed methodologies, and which aspects, according to them, are favored by using the automatic formative assessment with students. The teachers also had to explain the questions created for the formative assessment and the class of students chosen for testing the activities created. For each question, the teachers had to indicate:

- the dimension of the question (knowing, solving problems or justifying);
- the title of the question in the AAS;
- the main goal of the question;
- the topic of the question;
- the material from which they took inspiration for creating the question (from INVALSI, textbook, the internet, or if they invented the question);
- the strategies adopted to adapt the question to the automatic formative assessment.

After making the students carry out the activities they created, the teachers had to answer a final questionnaire. This questionnaire asked them to explain the activity carried out with the students. In particular, they had to describe:

- where the students carried out the activity (classroom, computer lab or at home);
- which students were involved (all students or only some of them);
- students' appreciation of the activity;
- difficulties reported to the students;
- changes made to the activities after the experimentation;
- the aspects favored by the use of the automatic formative assessment with students.

To carry out the analysis, a group of experts (the group of researchers who conducted the analysis) examined all the 51 questions and classified them according to the characteristics described by the teachers (dimension, topic, objective, etc.) and according to the completeness of the question, the number of sub-questions for adaptivity and the type of register required of students. For each category and sub-category, the various parts of each question and the types of response areas chosen were discussed, to reason about the possible difficulties teachers had (both in the transition from a question for standardized assessment to formative assessment and in the use of the AAS). For each dimension, the strategies used to adapt the formative assessment question were collected and analyzed. Then an example of a question was chosen for each dimension,

which was more exemplary of all the questions created by the teachers. Finally, the students' responses to the questions created by the teachers were analyzed to study the effectiveness of formative assessment and interactive feedback. For each question, we analyzed the following aspects:

- if students changed their answer in case of an error when multiple attempts were available;
- the students' most frequent mistakes;
- whether the guided process and interactive feedback were useful for the students to change their answers.

Therefore, for each question, the following data were analyzed:

- the number of students who answered the question: correctly, partially correctly, and incorrectly;
- the students' tests and their answers to the various parts of the question;
- any changes in the answers in subsequent attempts in case of incorrect answers.

Representative examples of this analysis will be shown.

The analysis of the students' activities was carried out from the researcher's point of view, and the results, not knowing the students, are assumptions based only on the answers given. Each teacher can carry out the same analysis by having much more information about students and obtaining even more significant results. Following the analysis, observations were made on the questions' design and on any changes that could improve it. The researchers' reflections were then compared with the teacher's reflections after carrying out the activity with the students, reported in the final questionnaire.

## 5 Results

### 5.1 General overview of the participants in the module and the activities carried out

17 teachers from the PP&S Community, one from the lower secondary school and all the others from the upper secondary school, took an active part in the course. 70% of teachers had already taken a course on using an AAS, and 65% had already used it with a class or group of students. The institute's typology in which the teachers teach was varied: Linguistic High School, Industrial Technical Institute, Scientific High School, lower secondary school, Technological-Electro-technical Institute, Classical High School. The teachers' subjects are: Mathematics (59%), Mathematics and Physics (24%), Mathematics and Computer Science (12%), Mathematics and Science (6%).

The teachers created three questions each for a total of 51 questions, one for each dimension. They used several sources to choose the standardized assessment question: an INVALSI question (37%), an exercise/problem found on a textbook (37%), an exercise/problem found on the internet (6%), an exercise/problem found in the Maple TA repository of the PP&S Project (10%). In the remaining 10% of cases, the teachers directly created a new question.

Regarding the dimension of “knowing,” the strategies they used to adapt the question to the formative assessment were:

- after a first closed-ended question by inserting an adaptive section in which, in the event of an incorrect answer, the student is guided in the resolution procedure;
- creating an algorithmic question so that every time there are different data;
- inserting final feedback that allows students who have made mistakes to understand them and to correct themselves;
- making an algorithmic multiple-choice question to bring out the most common misconceptions about a topic;
- inserting multiple areas of answers to understand different aspects.

In the dimension of “solving problems,” the strategies that the teachers used to adapt the question to the formative assessment were:

- asking questions with different registers (tracing a graph, filling in a table, and inserting a formula);
- contextualizing the problem in a real situation;
- making the question algorithmic;
- setting up computations to recognize the correct solution process in a multiple-choice question and adding a guided path in case of errors;
- guiding students in the resolution if they do not immediately respond correctly;
- using interactive feedback to help students who cannot solve the problem by guiding them step by step to the final solution.

In the “justifying” dimension, the strategies that teachers used to adapt the question to the formative assessment were:

- inserting after a closed-ended question a question with gaps in the text with missing words to choose from a list;
- allowing students to rephrase their justification after seeing the correct answer;
- developing a theoretical reasoning step by step;
- realization of a demonstration by inserting various multiple-choice questions;
- adding to an open answer the possibility of drawing the graph of the solution.

We asked the teachers how useful the various tools proposed were for the creation of the questions. Table 1 shows the teachers’ responses on a scale from 1= “not at all” to 5= “very much”.

**Table 1.** Tools used by teachers for creating questions [7].

| Questions  | Mean of teachers’ answers |
|--|---------------------------|
| Explanations followed during online meetings                 | 4.8                       |
| Notes taken during online meetings                           | 4.5                       |
| Videos of online tutoring                                    | 4.6                       |
| The material available on the platform on the use of the AAS | 3.9                       |

|   |     |
|---|-----|
| INVALSI website (www.invalsi.it)                | 3.9 |
| INVALSI tests archive (https://www.gestinv.it/) | 3.6 |
| Other materials found on the internet           | 3.0 |
| Tutor support via forums                        | 3.8 |
| Support from other teachers via forums          | 2.9 |

---

## 5.2 Examples of questions created by teachers

**Knowing dimension.** In the following example, the adaptation of questions designed for standardized assessment to formative assessment is carried out in several ways. To create the question, the teacher took inspiration from a problem found in the Project’s repository. The original question (Fig. 4) asks students to enter the line equation.

Find the equation of the line passing through the points:  $A=(3,-4)$  and  $B=(-1,-6)$ .



**Fig. 4.** Example of the question for standardized evaluation of the dimension “Knowing” [7].

Students can find the line equation in many different ways and using different representation registers (formulas, graphs, tables, etc.). At the end of the reasoning, students insert only the final equation they have found. In this case, an incorrect answer does not provide the teacher with precise information on the nature of the student’s difficulty. At the same time, the simple “wrong answer” feedback does not give students information about their mistakes and how to overcome them.

In the question created by the teacher, the request is split into two sub-questions, shown in Fig. 5. In the first part (on the left), the student must insert the equation of the line passing through two given points, however having two attempts available. By clicking on the “verify” button, the student can know if the answer entered is right or wrong. In case of a wrong answer, the student can reason again on the task and try to give a new answer. If the answer is correct, the question ends. In case of an incorrect answer, a two-step guided path is proposed to students to review the necessary theory and correctly answer the question. Next to the response area, students also have a button to preview the inserted answer and an equation editor to insert it. Moreover, the question has been made algorithmic, randomly varying the coordinates of the given points. In this way, students can try to answer it several times, and each time the points and the line passing through them will be different. Automatically students have many exercises with interactive and immediate feedback to practice.

**Solving problems dimension.** We propose two questions related to the “solving problems” dimension. The first one proposes a contextualized problem of solid geometry. For creating the question, the teacher took inspiration from an INVALSI item, shown

in Fig. 6. The question's main objective was to calculate the area and volume of the most common solid figures and give estimates of objects of daily life. Compared to the original question, the teacher added the total surface of the peeled jar as a second request and gave three attempts to answer (Fig. 7, left). The goal of the question has become to calculate the volume and the total surface of a cylinder. The strategy adopted to adapt the question to the automatic formative assessment was to make the question algorithmic and insert the request to calculate the total area. The question was created to guide the student in the resolution. After this first part, if students answer correctly, the question ends. If students make any mistakes, they are guided to solve the problem without knowing the correct answer. In the second part of the question (Fig. 7, right), in the two questions with drop-down menus, students can review the theory for calculating the volume of the can. After that, they have to enter the value of the base area and the volume. In the third and last part of the question (Fig. 8), students can review the theory for calculating the total area in a similar way to the previous one. In the second and third part of the question, if students click the "Verify" button, they have immediate feedback on their answer's correctness and on the correct answer (as shown in Fig. 8). In this way, students can try to correct themselves and rethink their reasoning before answering the question again.

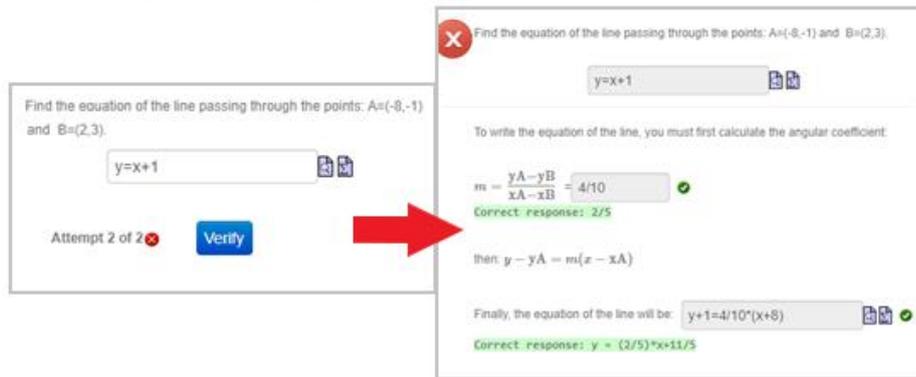


Fig. 5. Example of question of the "Knowing" dimension [7].

A peeled jar is 11 cm tall and the base diameter measures 6 cm.  
What is the volume of the peeled jar?



Find the correct answer:

- Around 100 cm<sup>3</sup>
- Around 200 cm<sup>3</sup>
- Around 300 cm<sup>3</sup>
- Around 400 cm<sup>3</sup>

Fig. 6. INVALSI question of solid geometry [7]. Source: GestInv 2.0 (www.gestinv.it).

A peeled jar is 12 cm tall and the base diameter measures 5.4 cm.  
What is the volume of the peeled jar?  
Find the correct answer:

- 274.6872000cm<sup>3</sup>
- 34.89060000cm<sup>3</sup>
- 101.7360000cm<sup>3</sup>
- 1098.7488cm<sup>3</sup>

What is the total surface area of the peeled jar?  
Answer:  cm<sup>2</sup>.

Attempt 2 of 3

---

To calculate the volume the formula is that of  (Click For List)

We must therefore calculate the value  that is:  cm<sup>2</sup>.

The value of the volume is:  cm<sup>3</sup>.

Attempt 1 of 1

Fig. 7. First and second part of the example question of the dimension “Solving problems” [7].

To calculate the total surface area the formula is

**Correct response:  $A1+2*Ab$**

The formula of lateral surface is:

**Correct response:  $C*h$**

The value of the circumference is:

**Correct response: 21.352 cm.**

The value of the total surface area is:

**Correct response: 392.8768000 cm<sup>2</sup>.**

Fig. 8. Third part of the example question of the dimension “Solving problems”.

The student can reflect on the correct answer and review the theoretical contents. Each phase of the procedure is characterized by immediate and interactive feedback so that the student remains focused on the task and is more motivated to move forward. This type of question helps students clarify what good performance is and offers opportunities to bridge the gap between current and desired performance. The question is algorithmic, so students can try to answer several times having new data. In this way, students can practice several times and consolidate a solving strategy.

The second example of the “Solving problems” dimension is of the “Data and forecast” scope. The goal of the question is to express the probability of compound events, starting from elementary events. The teacher was inspired by an INVALSI question shown in Fig. 9. The question asks to calculate the probability of a compound event contextualized. The strategy to turn the INVALSI question into a question for automatic formative assessment was to use interactive feedback to help students who cannot solve the problem, guiding them step-by-step to the final solution. The question has been made algorithmic: the number of coins of both types varies with each attempt.

The 1 euro coin and the Turkish 50 cents lira coin have the same dimensions, colors and measures. Mario has 3 one-euro coins and 2 Turkish fifty-cent lira coins in his pocket.



Without looking, he takes a coin out of his pocket and then another.

What is the probability that the first coin is fifty cents of a lira and the second coin is one euro?

- 15%
- 20%
- 25%
- 30%

**Fig. 9.** INVALSI question of probability. Source: GestInv 2.0 ([www.gestinv.it](http://www.gestinv.it)).

The first part of the question is similar to the original one, except for the numerical data which are algorithmic, and for the type of response area (numeric instead of multiple-choice). Students must answer by entering the probability value. They have only one attempt available. Students who answer correctly finish the question. Students who give the wrong answer have the error feedback but don't see the correct answer. A step-by-step guided procedure begins for them to find the correct solution (Fig. 10). In the logic of problem solving and computational thinking, the problem is broken down into three simpler problems: the probability of a first event, the probability of a second event, and then the probability of the intersection of the two events. To facilitate the students and guide them further in answering, the teacher has inserted a tree diagram that summarizes the possible events that can happen and has inserted a hint to remember that, at the time of the second draw, the total number of coins has decreased by one. In each part of the question, students have only one attempt, and, by clicking the "verify" button, they have immediate feedback on the correctness of the answer. If there is a mistake, they see what the correct answer was. In this way, they can rethink their reasoning, and they can correct themselves to try to do the rest of the question correctly. It is important to underline that, in the case of numeric answers, the answer can be entered in any form (such as fraction, approximate number, scientific notation), and, if correct, it is accepted in any case. As shown in Fig. 10, a student entered the answer in numeric form, and it was considered correct even though the correct answer is in fractional form. There is also an acceptability range of the answer to allow an approximation.

**Justifying dimension.** The following question proposes a problem concerning a floor's tessellation with tiles having different shapes represented by regular polygons. To create the question, the teacher took inspiration from a problem found on a textbook and inserted a step-by-step interactive process to lead students to the correct answer by following relevant steps. In the first part of the question, students must identify the regular polygon with the smallest number of sides that cannot be used for tessellation and enter the correct name in the text box. Then they must follow a guided path to justify the

correct answer (Fig. 11). The path for a guided demonstration is divided into three parts, in which the student must deal with different types of requests and different response areas (multiple-choice, numeric, choice from a list). Students have only one attempt available for each answer. The same path is shown to students. If the students answer incorrectly, they are shown the correct answer through interactive feedback. In this way, students can rethink the reasoning done and reason correctly on the next request. This type of question can be advantageous to train students to justify an answer by proposing a possible method. It can be useful to alternate this type of question with open-ended answers where students are asked to justify their answers freely.

The figure illustrates an interactive learning interface for a probability problem. It is divided into two main sections:

- Left Section (Problem Statement):**
  - Indicates an incorrect attempt with a red 'X' icon.
  - Text: "The 1 euro coin and the Turkish 50 cents lira coin have the same dimensions, colors and measures. Mario has 6 one euro coins and 3 Turkish fifty cents lira coins in his pocket." Below this are images of a 1 Euro coin and a 50 Cent Lira coin.
  - Text: "Without looking, he takes a coin out of his pocket and then another."
  - Question: "What is the probability that one coin is 1 euro and the other 50 cents of lira?"
  - Input field:
- Right Section (Guided Solution):**
  - Indicates a correct path with a green checkmark icon.
  - Text: "The following diagram can help you"
  - Tree Diagram:
    - Level 1: "The first coin is one euro" and "The first coin is 50 cents of lira"
    - Level 2 (from "one euro"): "The second coin is one euro" and "The second coin is 50 cents of lira"
    - Level 2 (from "50 cents of lira"): "The second coin is one euro" and "The second coin is 50 cents of lira"
  - Question 1: "What is the probability that the first coin is 1 euro and the second 50 cents of lira?"
    - Input:
    - Feedback: "Correct response: 1/4±0.05"
  - Question 2: "What is the probability that the first coin is 50 cents of lira and the second is 1 euro?"
    - Input:
    - Feedback: "Correct response: 1/4±0.05"
  - Question 3: "So what is the probability that one coin is 1 euro and the other 50 cents of lira?"
    - Input:
    - Feedback: "Correct response: 1/2±0.05"

Fig. 10. Second example question of the "Solving problems" dimension.

### 5.3 Observations on students' answers to questions created by teachers

**Solving problems dimension.** As a first example, the students' answers to the first question (Fig. 7 and 8) of the "Solving problems" dimension were analyzed. The question asked to measure the total surface and volume of the peeled jar through a step-by-step procedure. From the teacher's answers to the final questionnaire, we know that the students did the activity at home individually. We recall that this question was algorithmic, so each student answered a different question with the same structure but different values. A total of 20 students answered, of which 4 answered correctly, 6 answered incorrectly, and 10 answered partially correctly. Three students handed the test without giving answers, and three students tried the test but left the question blank. Then 13 tests were analyzed. Fig. 12 shows an example of how one student's answers appear in the gradebook. The answers given by the student are reported in the various attempts, indicating the correct and incorrect ones. The correct answers are shown. For the 3 students who answered the first part of the question correctly (both to the multiple-choice question on the volume and to the open answer on the total surface), the

question ended immediately. Another student also answered the first part of the question correctly after self-correcting at the second attempt. It is important to underline that, according to the settings chosen by the teacher, the students who missed the first part of the question, after clicking the verify button got a general error feedback. They did not know if the answer on volume, total area, or both were wrong. This may have confused some students. For example, a student gave an incorrect answer on the jar's volume on the first attempt, correct on the second attempt but again incorrect on the third. This happened because he left the second answer on the total area incomplete, and the system considers this as an error, so he was asked to try again. In all other cases, students who missed the first part of the question did not change the answer in subsequent attempts, maybe because they thought the answers were right, maybe because they did not know which of the two was wrong or how to change them. To offer a more precise feedback, it is possible to split the two questions into different parts, each one with a dedicated verify button. In this way, students would have accurate feedback on the correctness of the answer and know what they did wrong. Another possible modification to the question may be to specify the significant figures required in the answers (suggested only implicitly in the text of the first multiple-choice question). In 4 cases, the students received negative feedback even though the answer was correct but incorrectly approximated (e.g.,  $42.98 \text{ cm}^2$  instead of  $42.98660000 \text{ cm}^2$ ). The teacher can manually change the score assigned to students, but, in the case of numeric answers, it may be useful to add a tolerance range. Especially when the constant  $\pi$  is involved in the calculations, depending on the approximation used, the results change.

Figure 11 illustrates an example question of the "Justifying" dimension, showing four panels of a digital assessment interface. The top-left panel shows a question about tessellation: "Floor tiles can have different shapes, represented by regular polygons. What is the regular polygon with the smallest number of sides that cannot be used for tessellation?" The user has entered "pentagon" and the system shows "Correct response: pentagon". The top-right panel shows a question about the sum of internal angles of a pentagon: "To find the measure of an internal angle of a pentagon it is necessary to consider the sum of its internal angles and (Click For List)". The user has selected "divide by" and the system shows "Correct response: divide by". The bottom-left panel shows a multiple-choice question: "The justification for the correct answer starts from the reflection on the measurement of the internal angle of a pentagon. The sum of the internal angles of a pentagon is:" with options 720°, 180°, 360°, and 540°. The correct response is 540°. The bottom-right panel shows a question about the sum of three pentagons: "Since three pentagons must compete in a summit, there would be an empty space of:". A diagram shows three pentagons meeting at a vertex. The user has entered "30" and the system shows "Correct response: 36". Red arrows indicate the flow of the question parts.

**Fig. 11.** Example question of the "Justifying" dimension.

| Show all attempts <input checked="" type="checkbox"/>   |  |
|---|--|
| Your response   | Correct response   |
| A peeled jar is 8 cm tall and the base diameter measures 5.8 cm.  | A peeled jar is 8 cm tall and the base diameter measures 5.8 cm.   |
| What is the volume of the peeled jar?   | What is the volume of the peeled jar?  |
| Find the correct answer:  | Find the correct answer:   |
| <ul style="list-style-type: none"> <li><input checked="" type="radio"/> 211.2592000cm<sup>3</sup></li> <li><input checked="" type="radio"/> 211.2592000cm<sup>3</sup></li> <li><input checked="" type="radio"/> 211.2592000cm<sup>3</sup></li> </ul>  | 211.2592000cm <sup>3</sup>   |
| What is the total surface area of the peeled jar?   | What is the total surface area of the peeled jar?  |
| Answer:   | Answer: 198.5108000 cm <sup>2</sup> .  |
| <ul style="list-style-type: none"> <li><input checked="" type="radio"/> 188.7</li> <li><input checked="" type="radio"/> 191,6656</li> <li><input checked="" type="radio"/> 191,66 cm<sup>2</sup>.</li> </ul>  |   |
| $1.0 \times 1/2 + 1.0 \times 1/2 + 1.0 \times 1/2 + 0.0 \times 1/2 + 0.0 \times 1/2 + 0.0 \times 1/2 = 0.50$  |  |
| To calculate the volume the formula is that of<br><input checked="" type="radio"/> cylinder<br><input checked="" type="radio"/> We must therefore calculate the value dell'area di base that is: 26.4074 cm <sup>2</sup> .<br><input checked="" type="radio"/> The value of the volume is: 211.2592 cm <sup>3</sup> . | To calculate the volume the formula is that of cylinder<br>We must therefore calculate the value dell'area di base that is: 26.40740000 cm <sup>2</sup> .<br>The value of the volume is: 211.2592000 cm <sup>3</sup> . <div style="text-align: right;"></div> |

Fig. 12. Example of how the student's answers appear in the gradebook.

The guided procedure in the second and third part of the question was effective for half of the students. In some cases, the students who got the first part of the question wrong got it right in the second and third parts. Other students answered the theoretical questions correctly but answered the final question about volume and total area incorrectly. To understand if it is due to reasoning mistakes or inaccurate calculations, it may be useful to add a further sub-question for students who answer wrong, adding a step-by-step procedure to make the final calculation. This type of error may be due to problems in setting up and carrying out the calculations correctly or distraction problems. The teacher in the final questionnaire stated that, after the experimentation, she did not modify the question but invited the students to pay more attention.

As a second example of the "Solving problems" dimension, we analyzed the students' answers to the probability question (Fig. 10). 14 students answered this question. Only one student answered it correctly, 2 students did not answer correctly, and 11 students answered partially correctly. This question was algorithmic, so the question text was the same, but the number of coins of both types was different for each student. The students carried out the activity from home. For the only student who answered the first part of the question correctly, the question ended immediately. All the other students made mistakes of different kinds; in some cases, it is easy to understand the students' reasoning from the answer, while in other cases, it is more complicated. Students who have partially answered correctly can be classified into three groups: students who missed the first part of the question but did the others correctly (36%); those who got the first part wrong, did the second and third part correctly but got the last part wrong (36%), those who also made mistakes in the second or third part of the question (28%).

In the first part of the question, the most frequent mistake was considering only one event and computing the probability that the first coin was one euro, neglecting the second part. Another common mistake in the first part of the question was to insert a probability value greater than 1. In this case, it is not just a reasoning or calculation error. There is no control by students of the result obtained, either because they focused on the calculations and no longer considered the context of the problem or because they did not know the theory (the probability of an event is a value between 0 and 1). It may be useful to add theoretical feedback to remind students who make this type of mistake to make them reflect on their reasoning. For example, in one case only, a student entered a value greater than one (28, 4, 50, 3) in all four answer areas. A final kind of mistake was to misinterpret the data in the question. It had different numerical data: the number of coins of one type and the other, and then “1” referring to the one euro coin and “50” referring to the fifty-cent coin (repeated at the beginning of the text of the question and the end in the question). A student answered the first question:  $9/50$ . It seems that the number 50 (unjustified by the low numbers of coins in the text) was used by the student only because it is present twice in the text. It can be difficult for students to recognize which data they must use and which not. They tend to believe that they have to use all the numbers in the question text. For this reason, it might be useful to add to the step-by-step reasoning a section asking students to recognize and organize the given data.

The second and third part of the question was completed correctly by almost all students (72%) who, following the guided procedure, understood the mistake made in the first part of the question and self-corrected it in the subsequent parts. However, half of them got the final answer wrong. The error in the last part of the question may be due to a misunderstanding of the theory on the probability of the intersection of two events. The students responded by indicating the probability of a single event without considering the second event. It might be useful to add another part of the question with theoretical feedback to review the rule for students who made this kind of mistake. For greater effectiveness, the last item of the guided procedure could have been made explicit, for example, by highlighting the conjunction “and” or inserting the tree chart.

**Justifying dimension.** In this section, students’ answers to questions of the “justifying” dimension will be analyzed. As a first example, the answers to the question on the floor’s tessellation (Fig. 11) will be analyzed. The students carried out the activity at home, individually, and it was mandatory for the whole class. 16 students answered the question: 5 correctly and 11 partially correctly (none incorrectly). The students had immediate feedback on the correctness of each part of the question and the interactive feedback was proposed to everyone, even to those who correctly answered to the first part, to justify the answer given. If they made a mistake, they were shown the correct answer. In this way, they could rethink the reasoning made, try to correct themselves, and answer the following parts. Furthermore, the students had only one attempt to answer each sub-question. Consequently, the analysis focused on the type of error made by the students and on the effectiveness of the guided procedure to carry out the proof.

The 11 students who answered partially correctly made different mistakes in different parts of the question. Seven students (43%) answered the first question “What is the regular polygon with the smallest number of sides that cannot be used for tessellation?”

incorrectly: 4 answered “triangle”, one “square”, one “hexagon”, and one “none”. Of these, only two students, after seeing the correct answer, correctly completed all the other parts. Of the others, some have missed the second part of the question on the sum of the internal angles of a pentagon or the next part on the theory to derive the width of an internal angle of the pentagon. Examining the mistakes made in the second part of the question, it is interesting to observe that the students who incorrectly answered both to this and the first part all answered “360” while those who answered correctly to the first part answered “720”. When the question, like this one, is mainly theoretical, it might be useful to add as feedback a general rule for measuring the sum of a polygon’s internal angles. This is the essence of the interactive feedback: by inserting it in the form of a guided procedure, students will be forced to visualize it to proceed, and the feedback will be more effective. The mistakes that seem to be more difficult to interpret are those made by the students in the third part of the question, which was mistaken by 5 students. These students answered that to find the value of an internal angle of the pentagon, it was necessary to multiply (and not divide) the sum of its internal angles, and in the open answer, they wrote 180, 3, “ $x \leq 4$ ” (when asked, multiply by). The teacher will be able to deepen these answers later.

The step-by-step guided procedure seems to have been useful for the students. Even if they had made mistakes in the previous parts, almost all the students answered correctly to the final request on the angle that remains by putting three pentagons close together. This reasoning is composed of several parts: it is necessary to calculate the value of the pentagon’s internal angle, multiply it by three, and then subtract this value from 360. In this case, the wrong answers were 4, 45, 30, 180. It may help giving students multiple attempts to respond and, if they get it wrong, giving them the chance to self-correct. Alternatively, it might be useful to add another part of the question by dividing the requests into multiple response areas to identify the error in reasoning or the calculations. In the final questionnaire, the teacher stated that she would not change the question after experimenting with the students.

#### **5.4 Observations on the experimentation with students**

Analyzing the teachers’ responses to the questionnaire at the end of the experimentation, it emerged that students carried out the activity mainly in the computer lab of the school (47%), at home (35%), and in the classroom with mobile devices (18%). In almost all cases (94%), the students carried out the activity individually and not in groups, and it was mandatory for all students. The activity with students allowed the teachers to receive very useful feedback and reflect on the created tasks. Half of the teachers said that at the end of the experimentation, they would modify the proposed questions: by modifying the text of the question; adding explanations on how to answer; making the questions more accessible by proposing exercises for level groups (basic, intermediate and advanced); changing the methods of administration not as self-employment at home but as a compulsory classroom activity.

Regarding the classroom experimentation, the teachers evaluated various aspects shown in Table 2, on a scale from “1 = definitely no” to “4 = definitely yes”. All aspects have been assessed positively, and all scores are above three (except for the difficulties

encountered by the students). It turned out that the students appreciated the activity carried out, which was understandable for them, and that they appreciated the methodologies used, which also proved to be inclusive. It is significant and positive that the teachers are satisfied with the activity and that they want to propose it to a new class.

**Table 2.** Observations by teachers on different aspects of experimentation with students.

| Questions  | Mean of teachers' answers |
|--|---------------------------|
| The activity was understandable for the students             | 3.5                       |
| The students appreciated the new methodologies used          | 3.5                       |
| I would propose the activity to a new class                  | 3.5                       |
| The activity was inclusive                                   | 3.4                       |
| The students appreciated the activity                        | 3.3                       |
| The students were aware of the objectives of the activity    | 3.3                       |
| The students were aware of the prerequisites of the activity | 3.2                       |
| The students have achieved the objectives of the activity    | 3.1                       |
| I am satisfied with the activity                             | 3.1                       |
| Students with learning disabilities enjoyed the activity     | 3.0                       |
| The students had difficulties in carrying out the activity   | 2.5                       |

In some cases, the teachers said that students had some difficulty in carrying out the activity. They were of different types. Some students encountered technical difficulties in using the technology or using the AAS (because they used it for the first time or because they did not read the instructions correctly); others had trouble interpreting the requests correctly. Most of the teachers did not encounter difficulties in managing the activity with the students. Some teachers found it difficult to make students do the activity. According to one teacher, it was a mistake leaving the activity eligible because few students carried it out. The same problem was encountered by another teacher who regrets not having a computer lab available to carry out the activity, while at home only a few students finished it. A teacher pointed out that the teacher should always check the students' answers since different evaluations from the system's ones are often possible. He explained further: "if some students, having to provide the function of  $n$  instead of answering " $6n + 20$ ", answered " $6x + 20$ ", they clearly understood the meaning, then the answer cannot be considered completely wrong. However, at the same time, this makes me reflect on how important it is to be precise and exhaustive in writing the text of the question to the students". A few teachers, who were inexpert on the use of this AAS, found some difficulties in carrying out the experimentation and, in the end, realized that they would have made some changes in the organization or the questions. For this reason, they gave lower answers to some items of the questionnaire.

### 5.5 Teachers' observations on the training module

The teachers liked very much the contents and methodologies proposed within the training module. Table 3 shows their evaluations of different aspects of the module, on a scale of "1 = not at all" and "5 = very much". 88% of the teachers said they were satisfied with the training module and believed that it offered interesting ideas for educational activities. All teachers would recommend the attendance of this training module to a colleague.

**Table 3.** Satisfaction of teachers on the training module [7].

| Aspects  | Mean of teachers' answers |
|--|---------------------------|
| Clarity of explanations                                    | 4.9                       |
| Adequacy of the themes                                     | 4.8                       |
| Completeness of the justifications                         | 4.6                       |
| Method of conducting the course                            | 4.8                       |
| Opportunity to interact with the tutor                     | 4.9                       |
| Usefulness of the materials left available on the platform | 4.9                       |

Teachers were asked how much they believe that automatic formative assessment can favor students in different aspects reported in Table 4. The answer are given on a scale of 1 to 5, where 1 = "not at all" and 5 = "very much".

**Table 4.** Aspects promoted by the use of automatic formative assessment for students [7].

| Aspects   | Mean of teachers' answers |
|---|---------------------------|
| Review knowledge                                | 4.6                       |
| Understand the contents                         | 4.2                       |
| Develop problem solving strategies              | 4.4                       |
| Develop autonomy in problem solving             | 4.4                       |
| Develop justificatory skills                    | 4.1                       |
| Opportunity to practice for tests               | 4.6                       |
| Possibility to practice for the INVALSI tests   | 4.7                       |
| Facilitate study autonomy                       | 4.2                       |
| Promote metacognitive reflection                | 3.9                       |
| Promote student involvement in learning         | 4.1                       |
| Increase students' awareness of their abilities | 4.2                       |
| Understanding mistakes                          | 4.2                       |
| Increase motivation for the subject             | 3.8                       |
| Inclusion of students with learning disorders   | 4.1                       |

|   |     |
|---|-----|
| Student empowerment                       | 4.1 |
| Personalization of educational activities | 4.0 |

The values shown in Table 4 are all higher than 3.8. This means that teachers believe that automatic formative assessment can be an effective methodology for students, favoring multiple aspects. According to them, among the aspects most promoted by automatic formative assessment, there are the possibility of developing problem solving skills, reviewing the theoretical contents, and practicing for math and INVALSI tests.

Finally, teachers were asked how much they believe that automatic formative assessment can favor teachers with the different aspects reported in Table 5.

**Table 5.** Aspects promoted by the use of automatic formative assessment for teachers [7].

| Questions  | Mean of teachers' answers |
|--|---------------------------|
| Quality of teaching materials                            | 4.4                       |
| Greater attention to the activities proposed to students | 4.3                       |
| Professional development                                 | 4.5                       |
| Greater understanding of student difficulties            | 4.4                       |
| Control of student activities                            | 4.7                       |

Also in this case, the values shown in Table 5 are all higher than 4.3. This means that teachers believe that automatic formative assessment can be an effective methodology for teachers themselves. In particular, it can be useful to have quality feedback on the activities carried out by students, to understand their difficulties better, and to intervene with activities adapted to their needs. Further aspects identified by the teachers were:

- greater reflection by the teacher on the content of the questions and on the way they are proposed;
- make choices that respect the different learning styles of students, in consideration of the possibility of formulating questions of different types;
- promote a collective discussion and dialogue lessons;
- increase student engagement at home;
- organize didactic activities that facilitate the students' different learning styles.

When teachers were asked if they had encountered particular difficulties in customizing the course and building the activity, half of them answered "none". The other teachers reported, on the one hand, technical difficulties with the AAS (especially teachers who used it for the first time or with little consistency). Some more significant testimonies were: "I encountered some difficulties due to my lack of knowledge of the AAS, but by reviewing the videos of the course I was able to develop the questions"; "I have encountered the greatest difficulties in managing the time available: it is needed to try and try again to achieve a good result and this aspect has sometimes discouraged me. Now I am happy because I am finally beginning to acquire a bit of familiarity with the AAS"; "My difficulty lies in not being very familiar with AAS, for this reason, I have

built simple questions from a technical point of view.” On the other hand, difficulties of a methodological nature were reported, concerning the creation of questions for effective formative evaluation. For example: “I would certainly find it difficult to add a guided path in case of errors in the second question”; “In retrospect, I realized that I could have structured the questions differently to obtain an adequate evaluation”; “The greatest difficulty was finding suitable arguments for each dimension, for example, it is rather complicated to decide which question lends itself better to an argument. On the other hand, I believe that guided questions are particularly effective in “guiding” the argumentation, something that could hardly be done with a traditional test.”

## 6 Conclusions

In this paper, we presented a training module on automatic formative assessment for Mathematics teachers. Through an AAS, the teachers transformed questions designed for standardized assessment into questions for formative assessment, reasoning about the didactic context, the topic, and the application’s objective. Then they asked students to answer the questions and observed the activity carried out in the classroom. The teachers reflected on the aspects favored by the proposed methodologies for students and the teachers themselves. At the end of the module, the teachers shared the materials created with all the PP&S Community so that everyone could use and discuss them.

The participants appreciated the proposed contents and methodologies. They believed they had received enough tools to work independently with an AAS. The teachers were satisfied to have experienced the activities with their students, and they all would continue to use the automatic formative assessment during the school year. They could reflect on automatic formative assessment strategies, which helped different kinds of students learn Mathematics. They stated that using an AAS can have important advantages for students, such as developing skills, reviewing, and preparing for tests. There are also significant advantages for teachers, particularly for understanding students’ difficulties and needs and proposing adaptive teaching. The importance of immediate and interactive feedback in the proposed methodology is essential for both students and teachers. As shown in the results section, the AAS offers many tools for analyzing students’ activities. This helps us understand students’ behavior in answering a question, see if they self-corrected their mistakes in subsequent attempts, and understood the most frequent mistakes. These data can help us adjust the activities to adapt them to the students’ needs and to improve their usefulness for learning Mathematics. The examples discussed show the effectiveness of the immediate feedback and of the interactive feedback to guide students to the correct answer. Certainly, students may have technical difficulties using the AAS and in particular, in inserting the correct syntax. However, the rigidity of technological tools can educate them to read the instructions carefully. Furthermore, students must learn to use technologies also for educational purposes. Therefore, it is crucial to increase teachers’ training on the use of an AAS for formative assessment and train students to use this tool. This training activity presented in this paper could be further developed by collaborating with the teachers during the classroom activities. The researchers’ presence in the classroom would

support the teachers and increase the possibility of observing and studying the students' learning processes during automatic formative assessment activities. The use of automatic formative assessment supports adaptive teaching. Through the development of recent big data theory and learning analytics [29], we think it may be possible in the future to propose an adaptive educational system that uses data about students, learning processes, and learning products to provide an efficient, effective, and customized learning experience for students.

## References

1. Barana, A., Fioravera, M., Marchisio, M., Rabellino, S.: Adaptive Teaching Supported by ICTs to Reduce the School Failure in the Project "Scuola Dei Compiti." In: Proceedings of 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), pp. 432–437. IEEE (2017). <https://doi.org/10.1109/COMPSAC.2017.44>.
2. Shute, V.J., Zapata-Rivera, D.: Adaptive Technologies. ETS, Princeton, New Jersey (2007).
3. Bolondi, G., Branchetti, L., Giberti, C.: A quantitative methodology for analyzing the impact of the formulation of a mathematical item on students learning assessment. *Studies in Educational Evaluation* 58, 37–50 (2018). <https://doi.org/10.1016/j.stueduc.2018.05.002>.
4. Cascella, C., Giberti, C., Bolondi, G.: An analysis of Differential Item Functioning on INVALSI tests, designed to explore gender gap in mathematical tasks. *Studies in Educational Evaluation* 64, (2020).
5. Beatty, I.D., Gerace, W.J.: Technology-Enhanced Formative Assessment: A Research-Based Pedagogy for Teaching Science with Classroom Response Technology. *J Sci Educ Technol.* 18, 146–162 (2009). <https://doi.org/10.1007/s10956-008-9140-4>.
6. Barana, A., Conte, A., Fioravera, M., Marchisio, M., Rabellino, S.: A Model of Formative Automatic Assessment and Interactive Feedback for STEM. In: Proceedings of 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), pp. 1016–1025. IEEE, Tokyo, Japan (2018). <https://doi.org/10.1109/COMPSAC.2018.00178>.
7. Barana, A., Fissore, C., Marchisio, M.: From Standardized Assessment to Automatic Formative Assessment for Adaptive Teaching. In: Proceedings of the 12th International Conference on Computer Supported Education (CSEDU), pp. 285–296 (2020).
8. Borich, G.: Effective teaching methods. Pearson (2011).
9. Black, P., Wiliam, D.: Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability* 21(1), 5–31 (2009).
10. Sadler, D.R.: Formative assessment and the design of instructional systems. *Instructional Science.* 18(2), 119–144 (1989).
11. Osterlind, S.J.: Constructing test items. Springer, Dordrecht (1998).
12. van den Heuvel-Panhuizen, M., Becker, J.: Towards a Didactic Model for Assessment Design in Mathematics Education. In: Bishop, A.J., Clements, M.A., Keitel, C., Kilpatrick, J., and Leung, F.K.S. (eds.) *Second International Handbook of Mathematics Education*, pp. 689–716. Springer Netherlands, Dordrecht (2003).
13. Hattie, J., Timperley, H.: The Power of Feedback. *Review of Educational Research* 77(1), 81–112 (2007). <https://doi.org/10.3102/003465430298487>.
14. Pintrich, P.R., Zusho, A.: Student motivation and self-regulated learning in the college classroom. In: *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*. Springer, Dordrecht (2007).
15. Kearns, L.R.: Student Assessment in Online Learning: Challenges and Effective Practices. *MERLOT Journal of Online Learning and Teaching* 8(3), 198–208 (2012).

16. Khosravi, H.: Development of cost-effective adaptive educational systems via crowdsourcing. <https://er.educause.edu/blogs/2019/12/development-of-cost-effective-adaptive-educational-systems-via-crowdsourcing>, (2019).
17. Barana, A., Marchisio, M., Sacchet, M.: Advantages of Using Automatic Formative Assessment for Learning Mathematics. In: Draaijer, S., Joosten-ten Brinke, D., and Ras, E. (eds.) *Technology Enhanced Assessment*, pp. 180–198. Springer, Cham (2019).
18. Marchisio, M., Di Caro, L., Fioravera, M., Rabellino, S.: Towards Adaptive Systems for Automatic Formative Assessment in Virtual Learning Communities. In: *Proceedings of 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, pp. 1000–1005. IEEE, Tokyo, Japan (2018).
19. Barana, A., Marchisio, M., Rabellino, S.: Empowering Engagement through Automatic Formative Assessment. In: *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, pp. 216–225. IEEE, Milwaukee, WI, USA (2019).
20. Barana, A., Boffo, S., Gagliardi, F., Garuti, R., Marchisio, M.: Empowering Engagement in a Technology-Enhanced Learning Environment. In: Rehm, M., Saldien, J., and Manca, S. (eds.) *Project and Design Literacy as Cornerstones of Smart Education*, pp. 75–77. Springer Singapore (2020). [https://doi.org/10.1007/978-981-13-9652-6\\_7](https://doi.org/10.1007/978-981-13-9652-6_7).
21. Nicol, D.J., Macfarlane-Dick, D.: Formative assessment and self-regulated learning: a model and 7 principles of good feedback practice. *Studies in Higher Education* 31(2), 199–218 (2006).
22. Barana, A., Floris, F., Marchisio, M., Marello, C., Pulvirenti, M., Rabellino, S., Sacchet, M.: Adapting STEM Automated Assessment System to Enhance Language Skills. In: *Proceedings of the 15th International Scientific Conference eLearning and Software for Education*, pp. 403–410. Bucharest (2019). <https://doi.org/10.12753/2066-026X-19-126>.
23. Marello, C., Marchisio, M., Pulvirenti, M., Fissore, C.: Automatic assessment to enhance online dictionaries consultation skills. In: *proceedings of 16th International Conference on Cognition and Exploratory Learning in the Digital AGE (CELDA 2019)*, pp. 331–338. IADIS Press., Cagliari (2019).
24. Barana, A., Brancaccio, A., Esposito, M., Fioravera, M., Fissore, C., Marchisio, M., Pardini, C., Rabellino, S.: Online Asynchronous Collaboration for Enhancing Teacher Professional Knowledges and Competences. In: *The 14th International Scientific Conference eLearning and Software for Education*, pp. 167–175. ADLRO, Bucharest (2018).
25. Brancaccio, A., Marchisio, M., Palumbo, C., Pardini, C., Patrucco, A., Zich, R.: Problem Posing and Solving: Strategic Italian Key Action to Enhance Teaching and Learning Mathematics and Informatics in the High School. In: *Proceedings of 2015 IEEE 39th Annual Computer Software and Applications Conference*, pp. 845–850. IEEE (2015).
26. Fissore, C., Marchisio, M., Rabellino, S.: Secondary School Teacher Support and Training for Online Teaching During the COVID-19 Pandemic. (in press).
27. Bolondi, G., Ferretti, F., Santi, G.: National standardized tests database implemented as a research methodology in mathematics education. The case of algebraic powers. In: *Eleventh Congress of the European Society for Research in Mathematics Education*. Utrecht, Netherlands (2019).
28. Barana, A., Marchisio, M., Miori, R.: MATE-BOOSTER: Design of Tasks for Automatic Formative Assessment to Boost Mathematical Competence. (in press).
29. Barana, A., Conte, A., Fissore, C., Marchisio, M., Rabellino, S.: Learning Analytics to improve Formative Assessment strategies. *Journal of e-Learning and Knowledge Society* 15(3), 75–88 (2019). <https://doi.org/10.20368/1971-8829/1135057>.