

Effectiveness of Automatic Formative Assessment for learning Mathematics in Higher Education

Alice Barana, Marina Marchisio, Matteo Sacchet

Department of Molecular Biotechnology and Health Sciences, University of Turin, Italy.

Abstract

The health emergency due to COVID-19 has highlighted the need of new and flexible digital methodologies for learning and teaching Mathematics, which can support the individual student's needs and help shape education. In this paper, we propose to use Automatic Formative Assessment (AFA) activities designed according to an innovative model, framed on the theories of formative assessment and feedback. The model includes: availability of the activities and multiple attempts; algorithmic questions; open mathematical answers; contextualized tasks; immediate and interactive feedback. We analyzed an experience using AFA in a blended module in Mathematics for 96 first-year students in Biotechnology. We collected data from the platform on grades and attempts of AFA activities and cross-checked them with the final exam grades. The results show that the feedback gained from AFA helped students improve their performance and supported them in the exam preparation. In light of these results, we can conclude that similar activities can be an effective solution to support students' self-study during and after the pandemic.

Keywords: *Automatic assessment; digital learning environment; formative assessment; higher education; interactive feedback; mathematics education.*

1. Introduction

Higher education didactics has been affected by the ongoing health emergency that we are living in these days, like every other aspect of our lives has. Since March 2020, most of the lessons have been moved online, and it is necessary to find innovative and flexible solutions to support students, which could be adopted also in a “new normal” situation, in order not to create discontinuity in the teaching models (Aristovnik et al., 2020). Learning technologies can be very helpful, especially when they support interactive and adaptive activities which can be tailored on the individual learning paths. It is essential to use formative assessment to monitor and activate students as owners of their learning, allowing them to self-assess their achievements. In this paper, we propose to adopt a model of Automatic Formative Assessment (AFA) activities for Mathematics, largely used in secondary schools (Barana, Conte, et al., 2019), and we adapt it to higher education. After introducing the model in the theoretical framework, we present an experience that makes use of AFA activities to support learning in a blended modality in a Mathematics module for first-year students of a scientific Bachelor degree. Analyzing and cross-checking data about the AFA activities and the results of the final exam, we show the effectiveness of the model for learning Mathematics.

2. Theoretical Framework

2.1. Formative Assessment and Feedback

For the construct of Formative Assessment (FA), we adopted the Black and Wiliam’s definition (Black & Wiliam, 2009), which states that “*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.*” Critical to this conception of FA is the collection of evidence, and the interpretation and use of the information gathered to act on learning. The mere collection of students’ answers, without using them to alter and tailor their learning path, is not to be considered as formative (Wiliam, 2006). Among the strategies of FA, the provision of feedback is undoubtedly the most distinctive one and the object of in-depth studies. In the literature, results on feedback efficacy on learning are controversial (Azevedo & Bernard, 1995); for instance, from an outstanding review on feedback (Kluger & DeNisi, 1996), it emerges that in more than one-third of the 607 analyzed cases, feedback interventions reduced performance. This means that great attention should be paid to the feedback’s design. Hattie and Timperley (2007) provided a model for constructing effective feedback. The purpose of feedback is to reduce the discrepancy between current and desired understanding. Effective feedback 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. Moreover, feedback can work at four levels: at the task level, giving information about the task correctness; at the process level, adding details about the main steps needed to accomplish the task; at the self-regulation level, activating metacognitive processes; and at the self-level, adding personal evaluations about the learner. While the literature shows that the self-level is not effective, or even dangerous (Kluger & DeNisi, 1996), it seems that the only task-level feedback alone is not enough: many studies show that elaborated feedback are more useful than corrective ones to improve learning (Shute, 2008; Timmers & Veldkamp, 2011). The great part of elaborated feedback models that the literature proposes is static: students have to read them carefully and compare them with their results. Some studies also show that, more often than expected, students do not read them at all, especially if they perceive the task as too complicated or if they do not receive the feedback timely (Timmers & Veldkamp, 2011). It is clear that if the learners do not process feedback, the latter lose all their potentialities (Sadler, 1989).

2.2. Automatic Formative Assessment for Mathematics

When digital technologies are used to analyze the students' answers and to return feedback, we can integrate formative assessment and automatic assessment. We define Automatic Formative Assessment as the use of formative assessment in a Digital Learning Environment (DLE) through the automatic elaboration of students' answers and provision of feedback, where FA is intended as in the Black and Wiliam's definition (Barana et al., in press). In Mathematics, AFA is widely used to engage and motivate learners in online and blended courses. In this research, we adopted Möbius Assessment (previously known as Maple T.A.) as Automatic Assessment System (AAS), which, relying on Maple's mathematical engine, can process open-ended answers that are expressed in different registers from a mathematical perspective, and establish if they are equivalent to the correct solutions. This allows us to test different and complex cognitive processes. Moreover, it is possible to write algorithms to generate random parameters, formulas, graphics and even animated plots. Lastly, the AAS supports adaptive capabilities, so that the next question part depends on the previous answer.

Using this AAS, we developed a model to design AFA activities to enhance mathematical understanding (Barana, Conte, et al., 2019). It is based on the following six features: availability of assignments, which can be attempted in a self-paced way, without limitations in data, duration, and number of attempts; algorithm-based questions and answers, where random values or objects in the question text, answers and feedback appear different from student to student and at every attempt; open mathematical answers, graded through the advanced computing capabilities of the system; immediate feedback, shown to the students while they are still focused on the task; contextualization of the tasks in the real-world or interesting applications so that they can be relevant to students as well as for the discipline; interactive feedback, which is a step-by-step interactive guide to the resolution which appears just after attempting to autonomously answer the task. The Interactive Feedback (IF) is the

most original feature of our model (Barana et al., in press). In the IF, sub-questions investigate about prerequisites, simpler tasks or other representations of the initial problem, in order to guide students to a possible way to tackle the task. At each step, if they give the wrong answer, the correct one is shown to be used in the following steps. Moreover, they earn partial credits for the correctness of their answers in the step-by-step process (usually up to 75% of the question grade). These grades offer teachers and students more precise information about the learner's competence in a particular area. In AFA activities developed through our model, students can try the initial problem on their own and, in case of wrong answer, they have one or two more attempts available, so they are invited to try again and focus more on the task. If they fail, an IF shows a possible approach to the solution. This approach can help them identify their mistakes or give a different idea to solve the problem. Then, they can try the assignment again finding a similar problem, but with different data, so they have to repeat the whole process autonomously. Thus, the IF is a kind of elaborated feedback which acts at the process and self-regulation level, particularly relevant in making students process the feedback and use the information gained to improve their understanding.

3. Research Questions and Methodology

In previous studies, we have shown the effectiveness of our model of AFA activities and IF for learning Mathematics at secondary level (Barana, et al., 2019; Barana & Marchisio, 2020). In this paper, we are interested in measuring the effectiveness of our conception of AFA for learning Mathematics at university level, where the same model can be used to design activities. In particular, we try to answer to the following research questions: (RQ1) To what extent is the feedback obtained through our model of AFA useful to bridge the gap between current and reference performance in Mathematics in a scientific Bachelor? (RQ2) To what extent is the use of AFA activities designed according to our model useful to support the learning of Mathematics and preparation to the exam in a scientific Bachelor?

To investigate these research questions, we have analyzed data from a Mathematics module for the first-year students in Biotechnology at the University of Turin (Italy). The module was held in blended modality in the 2019/2020 a.y. It was not affected by the health emergency since lessons ended in January 2020; lessons were held in a classroom-based modality and support materials were available in a DLE for self-paced study, mainly in the form of interactive worksheets or AFA assignments (Marchisio et al., 2020). At the end of the module, the students took an exam, composed of a written part on the topics covered during the course and an oral discussion of an essay about a contextualized problem chosen and developed by the students. The online activities were not considered for the final grade, which was expressed on a scale from 18 (passing grade) to 31 (highest grade with honors). We drew several data from the AAS: for each student, the number of attempts to each assignment; the grade earned at their first attempt and at their last attempt to the tests (in

percentage); the number of students who attempted each test. From these data we computed several statistics: the rate of attempts per students, dividing the total number of attempts by the number of students who attempted the test; for each student, the average increase in the grades between the first and the last attempt; for each student, the total number of attempts to all the tests and the number of tests they attempted more than once. Moreover, we collected the exam grades and cross-checked them with data from AFA. We also analysed the answers to the questionnaire related to AFA. We used SPSS 26 for the analysis.

4. Results

Our analysis involved 96 students, of which 90 used the AFA activities and 73 took the final exam. 6 students took the exam without using any online assignment, while 23 students used the online activities but did not take the exam during the last academic year (a.y.). We started by analyzing the assignments on the AAS. The online course included 17 online tests designed according to our model. The assignments were attempted by a number of students ranging from 15 to 72 (average: 49.35). The total number of attempts per single test ranges from 29 to 453 (average: 184.94). The ratio of attempts per student ranges from 1.93 to 10.56 (average: 3.94). The last value was computed considering, for each assignment, only those students who made at least one attempt, since assignments were not mandatory. As a second step, we selected the students who made at least two attempts to one assignment (N=73) and we considered the average grade each student earned in their first assignment attempt and the average grade in their last attempt. Then, we compared the couple of variables through pairwise Student's t-tests. We found that students improved their grades by 12.27 points out of 100, and the increase is statistically significant ($p < 0.001$). This value gives us an idea of the sharp increase of students' scores when they are engaged in AFA activities aimed at improving their understanding. In order to qualitatively appreciate this improvement, we show how one student's answers changed during repeated attempts to a question with interactive feedback (Figure 1).

In order to understand the value of AFA activities to support the students' preparation for the exam, we compared the final exam grade of the students who used the online tests with those who never logged in the AAS through ANOVA. We found that students of the first group (N=5) had a lower grade than the others (N=68): the average grades are respectively 22.00 (Standard Deviation: 1.41) for the first group and 25.87 (SD: 3.54) for the second one. The difference is statistically significant ($p = 0.018$). Moreover, students who took the final exam used automatic assessment much more than students who did not take the exam; in particular, the first group made on average 38.60 attempts (SD: 45.94) to all the tests while the latter only 8.17 (SD: 16.65). The difference is statistically significant to the ANOVA test ($p = 0.003$). These results show that AFA activities were used to prepare for the exam, and that they were effective for this aim.

First attempt on 11th February 2020 at 5:31 PM

During a summer week, the maximum temperatures recorded at a seaside location over four consecutive days were 31 °C, 34 °C, 33 °C, 30 °C. The temperatures are measured in Celsius degrees.

What maximum temperature should be recorded on the fifth day, for the **mode** of the five readings to be 34 °C? ✓ °C

What maximum temperature should be recorded on the fifth day, for the **median** of the five readings to be 32 °C? ✗ °C

Step by step procedure when wrong: First step

Let's go step by step. According to what is known from the theory, the mode is the **most frequent** value of the distribution; furthermore, we have values for the first four days 31 °C, 34 °C, 33 °C, 30 °C, all different.

Since the value 34 °C coincides with one of the previous ones, the condition that the mode is 34 °C is equivalent to a maximum temperature of ✓ °C on the fifth day, so that this data appears exactly ✗ **Correct response: 2** times.

Step by step procedure when wrong: Second step

Since the number of data including the unknown is **odd** the median is given by **the central value** 32 °C should be the median with respect to the values 31 °C, 34 °C, 33 °C, 30 °C. Since the value is bigger than ✗ **Correct response: 2** of the other values and smaller than ✗ **Correct response: 2** of the other values, the maximum temperature of the fifth day will necessarily have to **coincide with** 32.

Step by step procedure when wrong: Third step

Indeed, if the maximum temperature of the fifth day f_5 is less than 32 °C, then there will be exactly ✗ **Correct response: 3** values between 31 °C, 34 °C, 33 °C, 30 °C, f_5 less than 32 °C, and this last value will not be the median. On the other side, if the maximum temperature of the fifth day f_5 is greater than 32 °C, then there will be exactly ✗ **Correct response: 3** 34 °C, 33 °C, 30 °C, f_5 greater than 32 °C, and this last value will not be the median.

Second attempt on 11th February 2020 at 5:53 PM

During a summer week, the maximum temperatures recorded at a seaside location over four consecutive days were 28 °C, 31 °C, 30 °C, 27 °C. The temperatures are measured in Celsius degrees.

What maximum temperature should be recorded on the fifth day, for the **mode** of the five readings to be 28 °C? ✓

What maximum temperature should be recorded on the fifth day, for the **median** of the five readings to be 29 °C? ✓

Figure 1. Two subsequent attempts to one question about Statistics by the same student. He failed the first attempt and followed the interactive feedback, composed of 3 steps, which guided him in the solving process. Then, he tried the task again finding different values; he solved it autonomously and gave the correct answer.

To better understand the relation between the use of AFA activities and the exam results, we looked for a correlation between the number of attempts to all the assignments and the exam grade. However, these variables did not result to be correlated: the exam grades grow when the number of attempts grows from 0 to 50 (excellent students made on average 51 attempts to the online tests), while for an upper number of attempts, the exam grades average decreases. We believe that it is due to the fact that excellent students had a lower need to train for the exam, while students with more difficulties in Mathematics needed to repeat each assignment more times to understand the underlying concepts and to manage to solve problems. Thus, we looked for another variable which could explain the exam grades, and reveal the students' commitment with AFA activities without being affected by the individual amount of work each student needed in order to succeed. We came up with the number of assignments each student tried at least once. This variable does not focus on the number of attempts made for assignment, but reflects the number of tests on which students made some effort. In so doing, it measures the level of engagement generated by the AFA. This resulted to be positively correlated to the exam grades (N=73, squared-R=0.32, p=0.008). In particular, we can observe how the exam grades increase when this variable increases: data are shown in Table 2. These results show that the feedback obtained from the assignments designed according to the model of AFA activities, when used to bridge the gap between current and reference performance, helped acquire useful mathematical knowledge and competence, as demonstrated in the final exam.

Table 2. Average of the exam grades and increase in the assignments grades during multiple attempts depending on the number of repeated AFA assignments.

N. of repeated assignments	N	Exam grade (average)	Exam grade (SD)
0	5	22.80	2.95
1-4	16	24.75	3.26
5-9	17	25.94	2.88
10-17	30	26.93	3.77

5. Conclusions

Through the data analysis and the results, it is possible to answer to the research questions. As regards (RQ1), it is possible to affirm that the feedback obtained through AFA activities helped students of a scientific Bachelor degree improve their performance in Mathematics. The feedback activated the students who generally repeated the tests to improve their performance; their average grades improved by 12.27 points out of 100 from the first to the last attempt. As regards (RQ2), the results show that the use of AFA activities was useful to support students' learning of Mathematics and their preparation to the final exam. It was observed that students used the online assignments to study for the exam and that the students who used the AFA activities earned a higher exam grade than those who never accessed them. Moreover, there is a significant correlation between the number of assignments that students repeated more than once and the exam grade. This variable measures the students' commitment and persistence in AFA activities, reducing the effects of the variability in the number of attempts to all the assignments due to the students' confidence, skills and need of repeating the activities many times to achieve good results. These results show that the model for designing AFA activities for Mathematics can be an effective methodology to support teaching and learning of Mathematics in blended and online modality in higher education. Probably, the feedback provided through these activities ensured that students with difficulties in Mathematics engaged in the tasks and better understood the solving process of the various problems, as well as the underlying mathematical concepts: in fact, the students who made the highest number of attempts were not the most brilliant ones. Thus, the AFA model can be a valid support for the weakest students even during the pandemic, since they might be the most damaged by the situation. These results cannot be generalized due to the low numbers included in this study. Moreover, the module adopted also other methodologies (such as problem solving) which could have influenced the results. As a further research it would be interesting to repeat similar analyses including higher numbers of students, even from different degree courses, and adding an entry test to check students' prior knowledge in Mathematics. The AFA materials developed for this module can be shared with other teachers in similar contexts and adopted on a larger scale, even in open online courses where students

are more autonomous but need timely feedback about their achievements. Further development of this study concerns Learning Analytics, developing solutions able to automatically detect problems or difficulties and help teachers make decisions to dynamically shape the path.

References

- Aristovnik, A., Keržič, D., Ravšelj, D., Tomaževič, N., & Umek, L. (2020). Impacts of the COVID-19 Pandemic on Life of Higher Education Students: A Global Perspective. *Sustainability*, *12*(20), 8438.
- Azevedo, R., & Bernard, R. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research*, *13*(2), 111–127.
- Barana, A., Conte, A., Fissore, C., Marchisio, M., & Rabellino, S. (2019). Learning Analytics to improve Formative Assessment strategies. *Journal of e-Learning and Knowledge Society*, *15*(3), 75–88.
- Barana, A., & Marchisio, M. (2020). An interactive learning environment to empower engagement in Mathematics. *IXD&A*, *45*, 302–321.
- Barana, A., Marchisio, M., & Sacchet, M. (in press). *Interactive Feedback for Learning Mathematics in a Digital Learning Environment*.
- Barana, A., Marchisio, M., & Sacchet, M. (2019). Advantages of Using Automatic Formative Assessment for Learning Mathematics. In S. Draaijer, D. Joosten-ten Brinke, & E. Ras (Eds.), *Technology Enhanced Assessment* (Vol. 1014, pp. 180–198). Springer.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Rev. Educ. Res.* , *77*(1), 81–112.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284.
- Marchisio, M., Remogna, S., Roman, F., & Sacchet, M. (2020). Teaching Mathematics in Scientific Bachelor Degrees Using a Blended Approach. *Proceedings of 2020 IEEE 44th Annual Computers, Software, and Applications Conference, COMPSAC 2020*, 190–195.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144.
- Shute, V. J. (2008). Focus on Formative Feedback. *Rev. Educ. Res.* , *78*(1), 153–189.
- Timmers, C., & Veldkamp, B. (2011). Attention paid to feedback provided by a computer-based assessment for learning on information literacy. *Computers & Education*, *56*(3), 923–930.
- Wiliam, D. (2006). Formative Assessment: Getting the Focus Right. *Educational Assessment*, *11*(3–4), 283–289.