

# FORMATIVE ASSESSMENT REVISITED: COMBINED USE OF MOODLE AND MAHARA FOR IMPROVING TEACHING GOALS IN SOME MEDICAL SCHOOL COURSES

Ana Sofia Lopes Sucuma<sup>1</sup>, Jacopo Garlasco<sup>2</sup> and Adriano Ceccarelli<sup>3</sup>

<sup>1</sup> Medicine and Surgery Course - University of Turin  
*ana.lopessucuma@edu.unito.it*

<sup>2</sup> Department of Public Health Sciences and Pediatrics - University of Turin  
*jacopo.garlasco@unito.it*

<sup>3</sup> Department of Clinical and Biological Sciences - University of Turin  
*adriano.ceccarelli@unito.it*

— FULL PAPER —

*TOPIC: Tertiary Education - Formative assessment, Continuous assessment - Portfolio - Ad*

## Abstract

Introductory courses employ mainly a summative style of teaching that fails to develop students' basic professional competencies useful in their future careers. Formative assessment, although time consuming, offers a greater opportunity for assessing deeper knowledge. We designed a course that employs continuous formative assessment, that relies heavily on peer feedback and uses these assessments as the basis for a final summative assessment. We demonstrate that this teaching strategy proves superior to the traditional style since it succeeds in attaining a deeper level of learning focused on competences. We also show that the ePortfolio platform Mahara is a notable addition to Moodle that allows students to display their work after profiting from peer assessment on Moodle.

**Keywords** – Moodle, Mahara, continuous assessment, formative assessment, feedback.

## 1 INTRODUCTION

All degrees at the Schools of Medicine of Italian Universities comprise both basic introductory courses and clinical courses. Teaching the latter aims at achieving the professional profiles of each degree, while the former are only regarded as introductory. They are designed to teach students theoretical notions necessary to master practical aspects in later clinical courses.

With this approach these introductory courses fail to contribute to the development of the key competences needed in each professional profile despite this being the main goal of tertiary education. Besides, the sense of continuous delay in the attainment of these skills can decrease the student's extrinsic motivation for learning and leaves intrinsic motivations as the single pulling force [2]. Moreover, the assessment methodology focuses mainly on obtaining and memorizing factual knowledge and not on the ability to apply those notions for more authentic problem solving tasks. The resulting levels of learning are thus only rarely above the comprehension level of Bloom's Taxonomy [1]. If learning has to be taken to a deeper level in courses normally regarded as introductory, these should aim at developing the "competence to develop one's own competences" together with more course-specific competences. Any delay in achieving these skills are wasteful and costly to students' training.

The first problem that emerges is how to introduce authentic assessment in these introductory courses.

Typically, these courses are short, normally within one term, and the number of students is high (on average >80). These features are in conflict with what would be needed to develop competences, namely continuous assessment combined with formative assessment. This way learning is obtained from a trial-error-assess-revise-trial cyclic process.

Formative assessment is activated by providing information on the learning process that teachers can use for instructional decisions and students can use for improving their performance. The latter motivates students [2]. This place the regulative aspect of feedback in a central position and assessments can be formative only if that feedback is used to modify the actions that will follow.

Formative assessment can be targeted to memorizing notions as well as achieving deeper learning, but the former is reduced to checking out correct answers to quizzes, and the little feedback that can be provided is limited to answering the same quiz for a second time and correcting the mistake. These types of tests are normally regarded as self-assessment formative tests but they only assess the ability to point out the correct answer. This will easily result in the student learning only which is the correct answer without necessarily having developed the skill to solve the quiz because that would actually require understanding what makes the correct answer correct. Nonetheless, a positive aspect of this style is that grading and feedback can be completely automatic with major savings in the time dedicated by teachers to the assessment activity.

Differently structured text such as short essays, essays, problem solving, etc are more conveniently used for assessing deeper learning and can be used for formative assessment [3]. The amount of time needed to carry out the tests and release feedback suitable to make the assessment formative increases remarkably with high numbers of students: in a 100 students class, assuming 30' as the average time to assess a 600 words essay, each test would require 50 hours of work per teacher, i.e. approximately an entire week. To understand the implications of this figure it should be considered that in the courses used to generate the data presented in this paper, the length of the modules ranged from 7 to 10 weeks and each was divided into 4 didactic units. Since the assessments took place at the end of each unit, the week required for the assessment meant that the feedback was made available while the next unit and therefore topic was well under way, and hence mostly useless for formative purposes. In summary, due to the very short nature of preparatory courses (or modules therein) when formative assessment of deeper learning is undertaken each test can only generate feedback on itself and the fast progression through the course's topics will impinge on the possibility to use the feedback for future tasks, hence decreasing the real formative potential of this approach.

In addition, both types of formative assessment (open-ended and close-ended tests) are only a transient phase in the teaching activity and are not by definition related to the final summative assessment. So in spite of the big effort that is required there is no guarantee that the effects of the feedback released during formative assessment will be reflected in an improved performance at final exam.

The second problem that emerges is the efficacy of formative assessment in short courses.

Assessing competences is also not immediately obvious in a final exam of the traditional type. On one hand it should be possible to assess the "ability to transfer one's knowledge and skills to solve a problem in a new and original context" where the context should be as similar as possible to a standard professional context where the problem could present itself [4]. On the other hand, a traditional exam is creating a completely artificial context that will never be found in a real situation in professional life. The student is deprived of any of the tools that are normally used by most professionals to approach problems: we first identify what we know, what we do not know, and what knowledge is needed to solve the problem. In the following step we search for the information we are missing through several channels, from textbooks, to scientific literature searches, as well as compare our view with colleagues and experts in the field, and then we will try to solve the problem.

The alternative to traditional exams, representing a one-shot picture of the student's performance, is known as continuous assessment. Continuous assessment is carried out during the whole duration of a course and aims at measuring the increase in competences with progress through the course [8]. In this way the student is assessed while a typical problem solving activity involving the expected competences is carried out in a normal context, i.e with all the tools needed to identify the missing information and search for it.

We have hypothesized that formative assessment tests that are required in each one of the taught modules can be turned into steps of a continuous assessment process while simultaneously making them as authentic as possible by allowing use of all the tools necessary for the performance required by the test.

In this paper we show that re-designing introductory modules according to this hypothesis results in a significant improvement in the student's final grades as well as in an increase in the student's perception of personal growth and deeper learning.

## 2 THE EXPERIMENT

In order to counteract these aspects that characterise the classical teaching style we designed a course whose main objective was that of combining formative assessment and authentic assessment into a continuous process by adopting the ePortfolio as an assessment tool and through the use of Moodle and Mahara.

The experiment included the Cell Biology course with 56 nursing students and the Genetics course with 104 medical students. During the academic year of 2019-20 both courses were held during the first term so that these two were not affected by the pandemic. Differently, during 2020-21 both courses were held entirely online due to COVID restrictions.

### 2.1 Academic year 2019-2020 A)

#### A. *Teaching tools*

Moodle was set up as a platform for individual work where students had access to all teaching materials, individual and group activities, forums for general discussion and shared virtual spaces for meetings.

#### B. *Course Design*

Students were split into 4-5 people groups (18 groups for medicine and 12 for nursing) and classes were held according to the following scheme:

1. Introduction: each unit had an initial opening lecture that introduced essential questions and key ideas on the topic of the current unit
2. Individual home study: At home and individually students made use of the teaching materials available on the Moodle platform for individual studying. These consisted of interactive video-lessons interspersed with quizzes.
3. In-class group work: in a second meeting, the class started with an end-of-unit (EOU) quiz containing questions similar in type and difficulty to the questions present on the Moodle interactive lessons. This was aimed at assessing their factual knowledge. After there was discussion of doubts and peer discussion of any problems that arose during individual studying. This was followed by group work tackling activities focused on problem solving.

At the end of each unit students had to present individual work in the form of short essays or written problem discussions related to a unit-specific topic. The best three out of four of the end-products as well as of the EOU quizzes results were used for final assessment calculated as simple mean of the scores.

At the end of the course students from the Medicine and Surgery group were also asked to fill in a questionnaire internal to the course with 52 questions regarding several aspects of the teaching-learning process. The questionnaire was organized into sections exploring the following aspects of the teaching/learning process:

1. Perception of own learning
2. Teaching by cognitive activation
3. Teaching by formative assessment
4. Perception of Molecular Genetics vs other modules/courses

## 5. EDUmeter questions

The questions in 5 are also present in the institutional tool used to measure student satisfaction at the University of Torino (Edumeter).

## 2.2 Academic year 2020-2021 A)

### A. *Teaching tools*

Moodle was set up as a platform for individual work where students had access to all teaching materials, individual and group activities, forums for general discussion and shared virtual spaces for meetings.

Differently from the previous year, a portfolio platform was added to Moodle together with continuous and formative assessment strategies.

Mahara was set up to harbour two separate institutions corresponding to the two groups of students (medicine, nursing). The members of one institution could see each other's profile, while they could not access the profiles of members of the other institution. On this platform students were asked to upload the products of the unit-specific assignments (artifacts). This represented the minimum number of artifacts that were necessary for final assessment. In addition, the students could upload any work they considered relevant evidence of their competencies. This required students to apply the critical skills needed to value one's own work and with multiple attempts sharpen that skill. Moreover, it encouraged peers to visit each other's pages to share opinions, criticisms and suggestions. This can be seen as an advantage since this feedback might extend beyond the limits imposed by a rubric.

### B. *Course Design*

Both courses were divided into 4 didactic units, and each unit had sets of interactive videos and self-assessment quizzes to test factual knowledge.

Students were split into 4-5 people groups (18 groups for medicine and 12 for nursing) and classes were held according to the following scheme:

1. Introduction: each unit had an initial opening lecture that introduced essential questions and key ideas on the topic of the current unit
2. Individual home study: students then made use of the teaching materials and exercises available on the Moodle platform for individual studying.
3. In-class group work: In a second meeting, the class met to discuss any doubts, work as a group, engage in peer discussion, and problem solving.

At the end of each unit students were asked to submit written group or individual work based on the assignment decided by the teachers. The work from each group/person was subsequently subjected to peer assessment through a Moodle workshop activity. Peer assessments were based on rubrics created by the teachers with the help of the students. After the feedback was released students were allowed to revise the group work individually, making changes where appropriate according to the suggestions. The revised work was then uploaded onto Mahara to be part of the student's portfolios and made available to the community of peers for feedback. Students could leave feedback till the end of the course. Importantly, two conditions were imposed: firstly, that the student keep the original copies of each piece in the portfolio and secondly, should they make any changes to the original, these should be explained critically and those explanations should be added to the portfolio.

The final assessment was based on students' self-assessment and teacher's assessment guided by a shared rubric. Then a meeting between the student and teacher took place to compare the assessment results from each side and through a discussion reach an agreement on the final score.

Differently to the previous academic year, the students from both courses were asked to fill in a questionnaire internal to the course. The questionnaire was organized into the same sections with the addition of a section devoted to the Perception of Portfolio Usefulness and aims.

### 3 DATA ANALYSIS

The distribution of final grades was analysed and the cohorts compared using the Dunn test for trends with Sidak correction [7].

Responses to the questionnaire from both years in the Medicine and Surgery course were analyzed with the Fisher's test [5].

### 4 EXPERIMENTAL RESULTS

#### 4.1 The combination of formative and continuous assessment improves the performance of students

We chose to compare the results of 2019 and 2020 as the content and organization of the courses were extremely similar, while they differed in the assessment. In the 2019 course at the beginning of each unit students were prompted with an "entry quiz" meant to assess the gaps in the knowledge of the basics, and could self-administer an identical copy of the entry quiz as many times as needed to self-assess while exploring and studying the content.

Final assessments were collected and scored for each student in each year, their distribution within each cohort were analyzed and compared for each year and each group of students.

Medians were calculated and compared, and the statistical significance of their difference was measured as described in the methods section.

We observed a significant ( $P=0,0001$ ) increase in the median value in 2020 courses compared to 2019.

Overall, 2020 results were higher than 2019 ( $p < 0.0001$ ) demonstrating an increasing trend (Figure 1). In addition to this almost no dispersion in the bottom quartile is observed for 2020 Medicine and Surgery course, while some dispersion is present in the data from 2019, indicating that it was not just an increase in the median value, but also a decrease in the dispersion of the bottom quartile, in which no scores  $< 18$  are present in 2020..

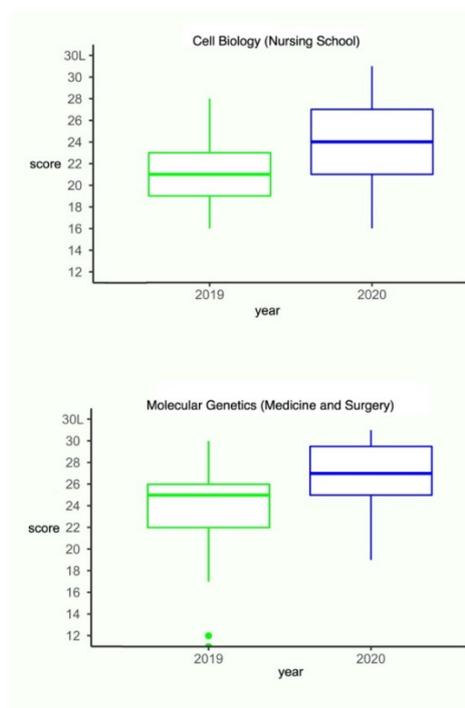


Figure 1. Distribution of final scores in Cell biology and Molecular Genetics Courses in academic years 2019 and 2020

The graph shows the position of the medians of each distribution with the 2nd and 3rd quartiles represented as boxes. 1st and 4th quartiles are indicated by lines. The outliers are indicated with coloured dots.

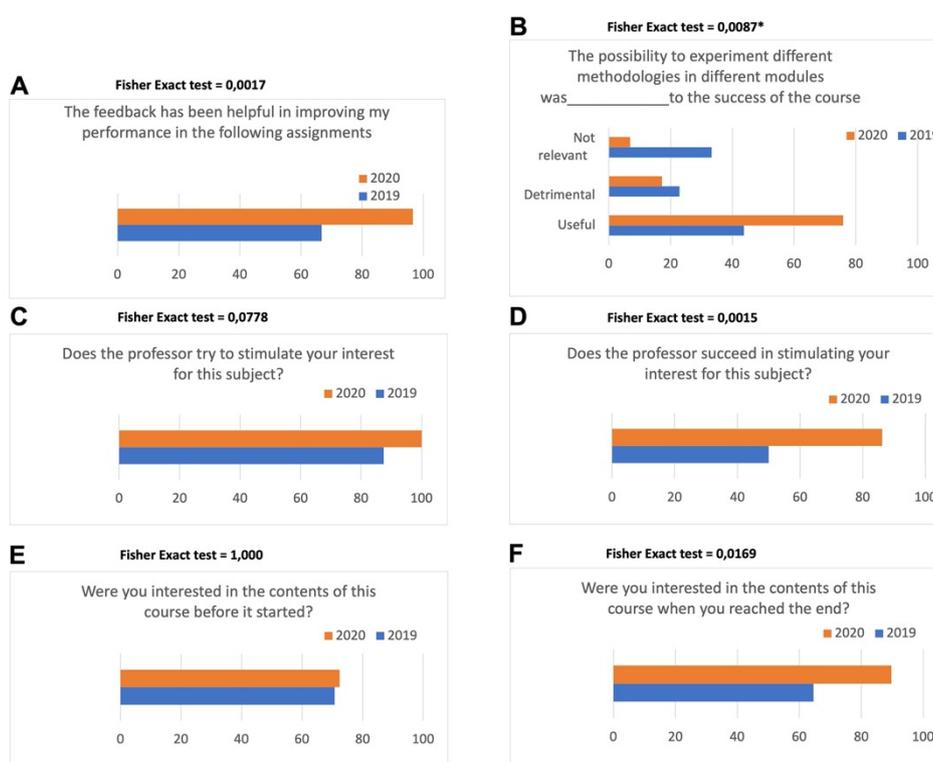
## 4.2 Combining formative and summative assessment by means of an ePortfolio improves the student's perception of their own progress and learning and promotes their motivation

Students responded to the introduction of the new teaching strategy with an improvement of the quality of their work, and we wanted to investigate their perception of the learning process and of the effectiveness of the adopted strategy.

We chose a set of 5 questions that looked into these aspects from the questionnaire administered in both years to one of the courses (Medicine and Surgery) (AC; manuscript in preparation). The questionnaire was not administered to nursing students during the first year so we can only present data regarding one out of two courses.

The students perceived the feedback as a positive factor in improving their performances (Figure 2 1, question A), and responded positively to the use of different teaching strategies in different modules of the same course (Figure 2, question B). During both years they perceived the effort of the teacher in trying to stimulate their interest (question C) but only in 2020 they regarded the teacher's activity as successful in stimulating their interest (question D). Consistent with this is the observation that their personal interest that they indicated to be present at the beginning of the course in both years had increased by the end of the course only in 2020 (questions E and F).

Taken in combination these results are supporting the notion that the new teaching strategy promotes self confidence as well as intrinsic motivations.



**Figure 2. Student's perception of learning and motivation**

Results from a questionnaire administered to the students of the Molecular Genetics course in academic years 2019 and 2020 are shown. The comparison was done between the frequencies of affirmative

answers, indicated on the x axis, the statistical significance was calculated with a Fisher Exact test and is indicated at the top of each graph.

The “\*\*” indicates that the comparison was made between the values of the choice “useful” that replaced the affirmative choice of the other questions

Years are indicated with different colours,

## 5 DISCUSSION

The continuous assessment strategy is usually regarded as one that puts more burden onto the teacher. This is due to the need to continuously assess student’s work. This is directly proportional to the size of classes and to the number of items to be assessed during the course, which in turn will determine the accuracy of the assessment. The bigger the class and the more accurate the assessment aimed for, the higher the work needed to assess and provide feedback on the student’s work.

We have found a way of incorporating the results of formative assessment, i.e. “usage of the feedback to alter future behaviour into the final product to be used for summative assessment. In this way we obtain two important results. First, we conveniently recycle the energy and the effort spent for formative assessment that in this way is directly generating the data to be used for the final summative assessment. Secondly, we can regard the link between the work carried out during the course and the final assessment with more confidence that the latter really assesses all the competences achieved during the course from the first day on.

We have used the Moodle workshop activity to guide students through the process of peer-assessment. Rubrics were made available in advance so that students could use them during the editing and elaboration of the artifacts [a][b]and then used for the assessment phase. This work is very conveniently done through this Moodle activity as far as setting the assessment criteria, allocating files to be assessed, and calculating scores with the included feedback. However, we think that the impossibility of making the assessed content as well as the feedback available to the whole community of a course after peer-assessment has taken place is a limitation. It would be extremely useful if all students could see the feedback released for each item, but to our knowledge the only available resource able to do this is the Moodle database that in turn presents other drawbacks. At the same time Moodle is not structured to obtain what we were aiming at: the collection of proofs of competences into a single container together with documents reporting metacognitive steps and critical thinking of one’s own work. If these had to be separate items in a database, it would be impossible to get the effect of assessing the work “as a whole” rather than summing up single scores into the final one.

We then decided to use Mahara, a platform widely known for building and storage of portfolios [6]. Here students could expose all their work to the entire community, exchange comments and edit it further when needed, or suggested by comments. Because we are aiming at assessing competences, we thought of the possibility of students presenting proof of their competences through different types of documentation, such as videos or picture galleries or presentations. However, although this was allowed and encouraged it didn’t occur in our first experience.

We also recognised that the combination of Moodle and Mahara worked best in a sequential manner. This permits students to give and receive peer feedback using a structured rubric and as part of a group initially and only then carry out individual self-assessment on Mahara. This ensures that the students have the time and the opportunity to develop the skill of critically assessing peer work and using that experience for their own self-improvement. Without this sequence students’ portfolios would suffer from lack of critical assessment and culminate into a collection of unscrutinized work.

The analysis of the results at the end of the two courses over two years, with introduction of the new strategy in the last year clearly shows an improvement in the final scores. The comparison of the distribution was carried out with the Dunn test, and the differences found are significant with a of  $P=0.0001$ . Another aspect we believe is important is that the difference in the confidence limits show an improvement also in the lower part of the curves, which in practical terms means that no students had negative results, and the minimum scores well far above the critical “sufficiency” threshold.

We think that our approach has allowed us to apply the formative assessment combined with continuous assessment strategies to courses that due to their size (>100 students) or duration (<30 hours) are commonly regarded as the most difficult situations.

We also appreciate that there is a need to investigate these results further and ascertain the role other factors might have played. A possible factor could be the change from in presence teaching to online teaching.

## BIBLIOGRAPHY

- [1] Bloom, Benjamin Samuel. 1979. *Taxonomy Of Educational Objectives*. London: Longman.
- [2] Cauley, Kathleen M., and James H. McMillan. "Formative Assessment Techniques to Support Student Motivation and Achievement." *The Clearing House* 83, no. 1 (2010): 1–6. <http://www.jstor.org/stable/20697885>.
- [3] "Educative Assessment: Designing Assessments To Inform And Improve Student Performance". 1999. *Choice Reviews Online* 36 (05): 36-2887-36-2887. doi:10.5860/choice.36-2887.
- [4] European Commission, Directorate-General for Employment, Social Affairs and Inclusion, (2019). *The European qualifications framework : supporting learning, work and cross-border mobility : 10th anniversary*, Publications Office. <https://data.europa.eu/doi/10.2767/750617>
- [5] Fisher, R. A. 1934. *Statistical Methods For Research Workers*. Edinburgh.
- [6] Mahara 1.6 user manual — Mahara 1.6 user manual [Internet]. [Manual.mahara.org](http://Manual.mahara.org). 2021 [cited 24 December 2021]. Available from: <https://manual.mahara.org/en/1.6/index.html>
- [7] Sidak, Zbynek. 1967. "Rectangular Confidence Regions For The Means Of Multivariate Normal Distributions". *Journal Of The American Statistical Association* 62 (318): 626. doi:10.2307/2283989.
- [8] Trotter, Eileen. 2006. "Student Perceptions Of Continuous Summative Assessment". *Assessment & Evaluation In Higher Education* 31 (5): 505-521. doi:10.1080/02602930600679506. Riferimento 1 [Arial 10pt, allineamento giustificato, maiuscolo e minuscolo]