Received: 1 February 2018 | Accepted: 19 March 2018
DOI: 10.1002/cae.21950

RESEARCH ARTICLE

WILEY

Feedback by automatic assessment systems used in mathematics homework in the engineering field

Jorge Gaona[1,2,3] | **Mercedes Reguant**[4] | **Ismael Valdivia**[3] | **Mario Vásquez**[3] | **Teresa Sancho-Vinuesa**[5]

[1] Paris Diderot University, Paris, France
[2] Laboratoire de didactique André Revuz (EA 4434), UA, UCP, UPD, UPEC, URN, Paris, France
[3] Technological University of Chile (INACAP), Huechuraba, Chile
[4] University of Barcelona, Barcelona, Spain
[5] Open University of Catalonia, Catalonia, Spain

Correspondence
Jorge Gaona, Paris Diderot University, 8 place Aurélie Nemours, Batîment Sophie Germain, Bureau 747, Paris, Ile de France, France 75013.
Email: jorge.gaona@etu.univ-paris-diderot.fr

Funding information
CIEDU Universidad Tecnológica de Chile INACAP, Grant number: Proyecto SEDOL-M

Abstract
This research presents the results of the teaching innovation *Dynamic Online Assessment System in Mathematics*, which is implemented in higher education to promote self-study by students outside the classroom. The WIRIS calculator was integrated into the Moodle platform to create questions with random elements, for example, students had access to different variants of the same question. The effect of the type of feedback (immediate or deferred) on the work of the students on the platform, measured by means of participation, time spent, and grades obtained, was evaluated. We used a quasi-experimental methodology for a population of 5,507 students, distributed in 229 courses on four campuses that learn Mathematics I in engineering programs. Immediate feedback exhibits better work of students on the platform, but this work is not necessarily more efficient in comparison with the work performed by students using assessments online assessment with deferred feedback.

KEYWORDS
automatic assessment tools, educational innovation, feedback, mathematics assessment, web-based homework

1 | INTRODUCTION

Innovation in education and a rigorous analysis of the results of experience are two complementary actions between each other and consubstantial to the improvement of educational systems. This is particularly true for professional mathematics subjects, where the academic results of the students can be improved.

Teaching and learning of mathematical content are of great concern for institutions and governments worldwide, with specific characteristics in Latin America, particularly in Chile. In institutions such as the Technological University of Chile (INACAP), this concern is particularly critical because of the academic and social impacts that this discipline has on the future professionals trained in this institution.

To improve the mathematical education in this institution, an educational innovation project has been promoted, with a focus on optimizing the self-regulatory processes of students, particularly in terms of time management and the development of study habits. Considering this, a dynamic assessment system with automatic feedback has been developed using the WIRIS QUIZZES tool. This tool is integrated into INACAP's *Learning Environment*, which is based on Moodle.

This project, referred to as Dynamic Online Assessment System in Mathematics (SEDOL-M by the Spanish acronym), has a general purpose of developing and implementing a system of questions with automatic feedback, with a focus on enabling self-study by the students of Mathematics I outside the classroom, at INACAP.

This institution has 120,000 students in 26 campuses that are distributed throughout Chile, of which approximately 40,000 students attend Mathematics I each year. Given the size of the institution and the volume of students in this field, the project has been gradually scaled up, this is because positive results are difficult to replicate with ordinary teachers on a small scale and in controlled environments [2].

To achieve an optimal scale-up, focus has been placed on studying the conditions that encourage active participation of students. Among the array of decisions for the design and implementation of such a project, the time when automatic feedback is delivered to students has been discussed, and in this article, we focus on analyzing the impact of work on students, according to the type of feedback used. For this, the data collected during the first semester of 2017 in four INACAP campuses is considered.

2 | PROJECT DESCRIPTION

SEDOL-M is a project promoted by INACAP's Center for Innovation in Education (CIEDU) from a small project developed by a group of professors from one of the INACAP campuses, funded by the CIEDU itself through an in-house contest in 2013.

In this first experience, the project was implemented with approximately 200 students; through statistical tests, it was concluded that there was a positive correlation between the use of the platform and the results of the students in a paper/pencil-based evaluation and external to the intervention; that is, an improvement was observed in the results of the students [11]. These results are in agreement with other interventions of this type in similar contexts [4,23], and from these significant results, the project was scaled up to the entire institution.

The design and implementation strategies are mainly based on the formation of design teams in each campus that programmed each question of the system and with whom the decisions regarding the implementation rules for both the teachers and students have been discussed. The project began in August 2015 with the design phase. The first pilot project was implemented between March 2016 and August 2016 with 1,200 students from four campuses. Between March 2017 and August 2017, the project was extended to 100% of these four campuses and a fifth campus was integrated, thereby reaching a total of 8,500 students distributed in 339 courses with 96 professors. For the March 2018–August 2018 period, the project aimed to cover 13,000 students from eight campuses. This study was conducted with a part of the population chosen according to the criteria described in the methodology during the March 2017–August 2017 semester.

From a technological point of view, the assessment system is developed in the Moodle platform, which integrates the WIRIS calculator (http://www.wiris.com) to create questions that contain random elements (numbers, symbols, and graphics) in a statement. This randomness in statements has two scenarios: one in which two students face the same type of question but with certain variations and the other in which each student answers the same questionnaire several times with different parameters.

A vast majority of the questions (over 95%) was designed using short-answer questions. The students were provided a blank space to answer these questions with the help of a mathematical editor or a handwriting recognition system.

This system is equipped with a computer algebra system (CAS). CAS allows us to compare the answer provided by a student with the answer pattern previously defined by the team of teachers, thereby enabling us to recognize mathematically equivalent answers, such as 0.5 and 1/2, x/2 + y/2 and (y + x)/2, or 1 + tan^2(x) and sec^2(x). It also allows the differentiation of certain features of mathematical objects, such as whether they are simplified or factored, among other properties.

An example of a question with the features described above is shown in Figure 1. In this question, the random elements are the color of the rectangle for which the surface area and perimeter are requested, the letters and the coefficients used for the expressions that define the magnitudes of the segments, and the size and position of the rectangles.

Once the student answers the question, the system not only immediately reports whether the answer is correct or incorrect but also provides a strategy for solving the task set according to the random parameters. For example, in the question shown in Figure 1, once the student has responded, the system provides the student step-by-step feedback, as shown in Figure 2.

As in all computer systems for educational purposes, in addition to the potentials described above, several limitations exist. For example, graphics are still images with which the student cannot interact. Moreover, similar to other systems of this type, only the final answer can be evaluated, instead of the work process of the students. Finally, the feedback—although it contains valuable information—does not completely replace the information and the personal presence offered by a teacher.

In the design and implementation of this project, the teaching staff has made a series of decisions of a different nature, some of which were specific to the disciplinary content, whereas the others were of a more generic nature.

Examples of specific decisions concerning the discipline include the types of tasks to be designed, the semiotic registers used in the questions (e.g., graphical, algebraic, and tabular), and the detail with which the feedback is given step by step. These options are limited because of the specific characteristics of WIRIS.

General decisions are those that impact evaluations as a whole. Examples of such decisions include the time (in min) given to students to answer an quiz, the time (in days) during which each quiz can be answered, the number of opportunities

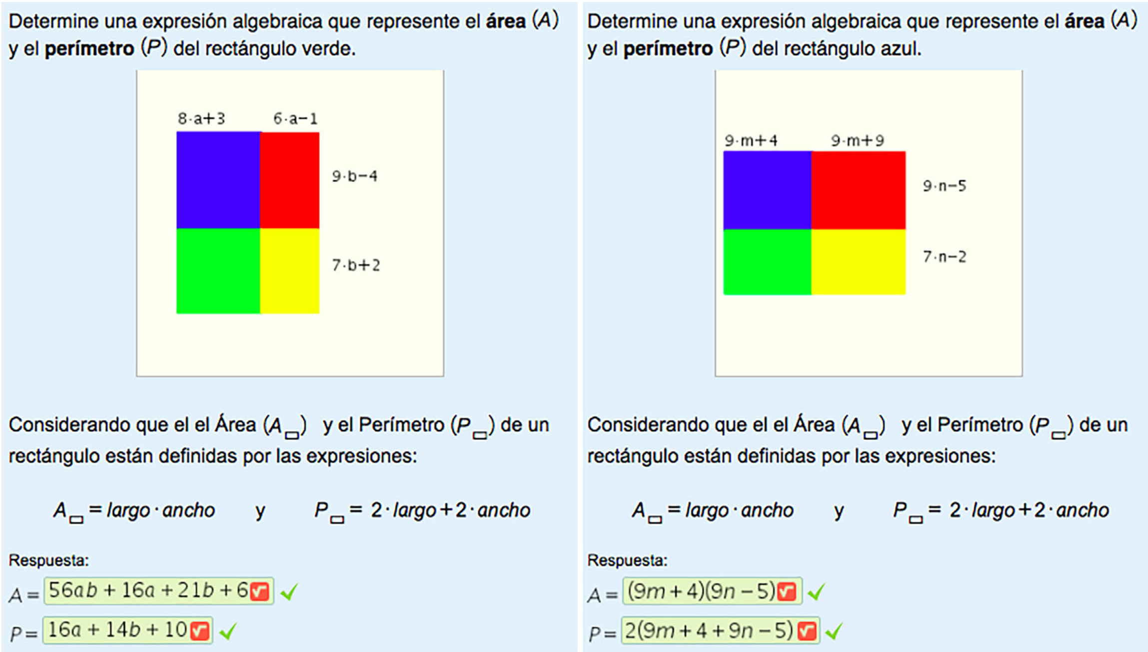

FIGURE 1 An example of an algebraic question with random parameters

they have to respond to the same quiz, the method of calculating the grade (highest or average grade), and the way in which the feedback is delivered (immediate or deferred).

Herein, the analysis of the way in which the feedback is delivered was further studied. In particular, the effects on the work of students in the platform were studied based on the type of feedback (i.e., deferred or immediate) with which the quizzes were configured. All other rules for students remained the same and were described in the methodology.

In immediate feedback, after each question, the student can click on the "check" button. When the button is clicked, the platform displays whether the answer is correct and provides a step-by-step feedback.

In deferred feedback, the student can click on the "send everything and finish" button only once: at the end of the quiz. When the button is clicked, the platform displays whether each answer is correct and provides a step-by-step feedback.

During the pilot project of 2016, only immediate feedback was used; however, the teachers noticed that the students

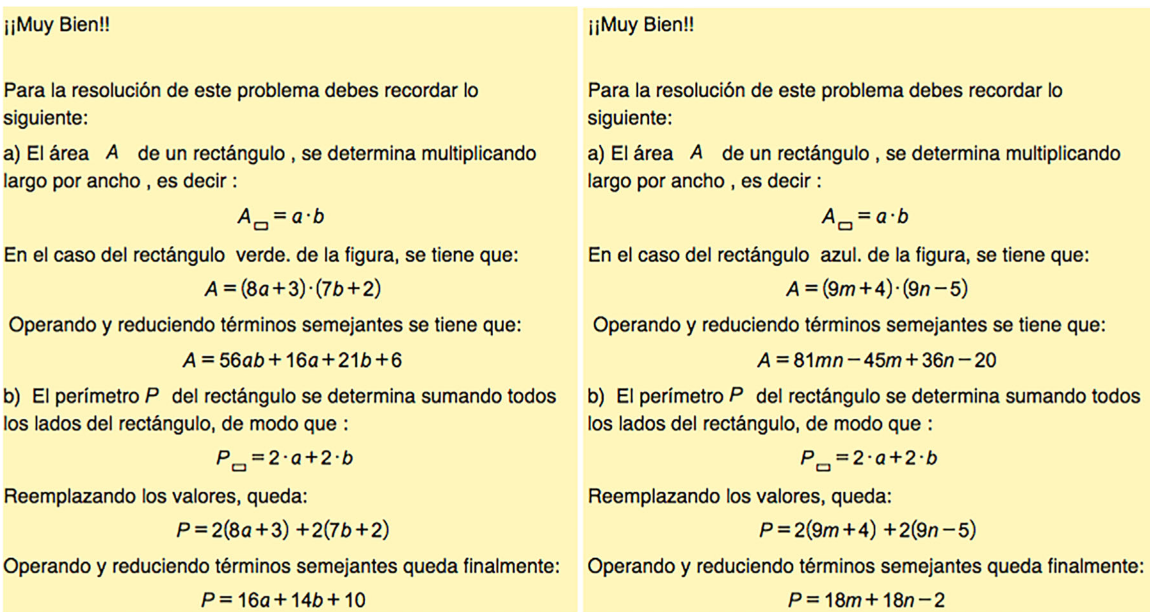

FIGURE 2 Step-by-step feedback based on the random parameters

started answering the questions. Whenever they entered an incorrect answer, they completed the questionnaire and started it again. The students could answer a questionnaire as many times as they wanted until the test was open, as detailed below. This led the teachers to question whether this was the best option. These doubts motivated the present study, which determines which type of automatic feedback contributes to a better work experience outside the classroom.

Herein, foundations are laid for the theoretical framework of the study, including the methodology of analysis and the most relevant results and their discussion, followed by the conclusions and future avenues of research.

3 | THEORETICAL FRAMEWORK

The theoretical approach of this research is based on three basic pillars. First, educational innovation, definition, and implementation of the project are explained as a response to the change in higher education, with the purpose of developing competencies, which is the expression of learning achievement. Second, the most relevant elements of a student's autonomous work are explained as the axis of the implemented tool. Third, an approximation to the state of the art is made to the evaluation and feedback, which act as consubstantial elements to the proposed innovation.

3.1 | Teaching innovation in higher education

Starting from the global scenario and the increasingly complex and changing world of today, we have the obligation to prepare professionals who can respond to these situations and will face challenges once they occupy their working space. In this regard, colleges face a great challenge because they need to prepare a highly qualified professional resource from a technical point of view and contribute to the training of a professionally competent, critical, and autonomous individual who can act on their own and seek appropriate solutions for the issues that arise [1].

As a reflection of the above aspects, different types of changes in higher education have been witnessed in the past 20 years. In this regard, we believe that the curricular mesh of the degree must be a coherent framework wherein all disciplines act interdependently, thereby contributing to the training required by a professional individual. Nevertheless, each discipline must evaluate its own disciplinary logic to achieve global goals. Consequently, each subject reviews its contents and strategies so that both elements converge to achieve the objectives.

Because of this reason, the results of the innovation presented herein acquire a special value since its purpose is to explain the increase in a student's studying time. This innovation also attempts to improve the results in terms of learning, while it becomes an educational device that favors learning autonomy.

3.2 | Web-based homework

In onsite teaching, the autonomous work of a student or the tasks outside the classroom—usually called *homework*—are a set of activities performed by the student based on a teacher's request to aid the student in understanding the concepts and procedures necessary to advance in a subject appropriately. This set of activities constitutes the student's educational evaluation [19, p. 24]. If these activities are performed over the Web (i.e., *web-based homework* [WBH]), they provide the necessary information to the teacher for assessing the learning experience, identifying particular difficulties, improving the learning process and the confidence of the student as well as their perception of learning [3,14]. However, it is also important to consider the immediacy with which the student receives a response without the teacher being burdened with additional work during the class or during consultation meetings. In brief, WBH systems offer higher education institutions (with limited resources) an option of providing students an individualized feedback on exercises and problems at a reasonable cost [16].

Several studies have analyzed the effectiveness of this work, particularly in elementary and secondary education. These studies have shown that a combination of these systems with teacher training can improve the results of learning mathematics [20].

In higher education, the most relevant studies on basic mathematic courses have analyzed the effectiveness of tasks outside the classroom through a web-based platform and associated it with the students' attitudes toward mathematical learning [17,19]. This shows that students' attitudes can be involved in the execution of the mathematical work proposed in these digital scenarios [6]. Other studies have argued that these positive results rely on the quality of the tasks and the proposed feedback, the outcomes of which are limited by the technical features of the software chosen as well as the instrumental dimensions linked to design [21,22].

However, some studies have not found significant differences in the results of the students according to the type of homework. The performance of students in statistics is not contingent upon the type of homework or its modality, but it does depend on the experience of the teacher and on the students' skills [19].

3.3 | Evaluation and feedback

Concerning evaluation, there remains a huge punitive burden in the definition of the concept. Despite all the advances at a theoretical level, there is still a commonplace or collective perception that the evaluation means a grade or a mark, that is, the evaluation is merely reduced to measuring performance.

In this regard, by referring to a study by Escudero [9] on a brief tour of the concept of evaluation, four major periods can be identified. However, in this study, we focused on the concept [12] at the end of the 80s. They state that evaluation is considered as a "generator of evaluative culture," which implies a continuous improvement of the processes based on the results and a change in the attitude toward evaluation.

The teaching innovation proposed by INACAP and our research finds their framework in this fourth generation of the evaluation, which is approached from two different perspectives. According to the first perspective, this research evaluates one of the variables introduced by the innovation tool to assess the improvement achieved; however, according to the second perspective, the evaluation is based on the type of the evaluation of learning proposed by the innovation tool as a resource for students, who will be able to regulate their learning (process).

With regard to this last aspect, the importance of evaluation for learning emerges in the first place, an element widely studied in the literature. From the perspective of students, evaluation provides information about the level of development of competencies, the achievement via learning, the adequacy of time and efficiency of the strategies used for the study, particularly in the case of an intrinsic evaluation designed with the teaching–learning process. Regardless of whether it is a diagnostic, educational, or summative evaluation, the evaluation itself has an educational quality. It guides students with respect to the topics emphasizing the studying effort and reports the achievements, as already stated in the previous subsection.

In this case, at least, the following conditions must be met: the evaluation must be feasible/sustainable, consistent with the expected learning results, and provide sufficient opportunities for evaluation as well as information in return in a timely and appropriate manner. A previous study stated that it is not about making a judgment at the end of the journey but about following the progression of the development of competences") [10].

At this point, we introduce the concept of feedback, which is understood from its more classical definition. This means that information is provided to students regarding their performance from a much more global and comprehensive perspective, that is, the information provided as part of a continuous process of support for learning improves the future performance. "Evaluation is conducted based on an educational approach, and evaluation itself is an opportunity to learn through feedback and corrective practice" [8]. A previous research proposed spiral feedback, an idea that we particularly are interested in because of its propositional approach, with the idea of improvement and transformation involved in this concept [5].

From this dynamic perspective, we can establish a good practice of feedback in relation to the factors that can strengthen the ability of students to self-regulate their learning. A good practice of educational feedback should facilitate the development of self-evaluation in learning as it provides students with high-quality information about their learning [18], promotes positive motivational beliefs and self-esteem, and provides information to teachers who can help shape teaching.

Nevertheless, some studies have warned about some limitations of the feedback provided in similar systems. For example, Cazes and his colleagues [6] performed a qualitative analysis and demonstrated that many times the feedbacks provided by these systems are mathematically unfit for the work done by the students or fit for the purpose but misinterpreted by the students or fit but insufficient. It is such that the system instead shows what students do not know more clearly than generating certainty in them and, therefore, the teacher's role as a mediator with this type of technology is essential [1].

In contrast, it is an aid in the learning process and grants the practice-feedback combination the category of formative evaluation [25]. Different studies can be found in literature on the use of online quizzes, which can be similar to online homework as in the case of Johnson and its impact on learning [15].

4 | METHODOLOGY

For this study, we followed a quasi-experimental research methodology with two equivalent groups and no control group in which successive measures have been conducted but always within intervention, that is, under the influence of the independent variable. The independent variable is the type of feedback from the tool, and the dependent variables are as follows: the participation rate of students, the number of attempts, duration of working with the platform, GPA, and maximum scores obtained using the platform, all of them per student.

4.1 | Purpose of research

The primary purpose of the research is to gather useful information to drive the decision-making process in terms of the type of feedback—deferred or immediate—which should be used in the next phases of implementation of SEDOL-M at INACAP. In particular, this study aimed to investigate the following aspects:

1. Discover the relationship between the type of feedback provided by the system—deferred or immediate—and its contribution to or insufficiency for the exercises offered by the tool. The latter is a desirable behavior that needs to be investigated.

2. Analyze the differences between the students who used the tool with immediate and deferred feedback, measured based on the number of attempts, work duration, GPA, and the average of the maximum grades.

4.2 | Population

As indicated in the Introduction section, INACAP has 26 campuses, covering courses ranging from technical training to university degrees, in different regions of the country. All first-year students choose Mathematics I, which was academically created as a subject that levels mathematical knowledge to be learned in secondary education [24]. It has different variants according to the professional degree in which it is taught. For example, the engineering degrees have a version of Mathematics I coupled with trigonometry and complex numbers, whereas in the field of design, Mathematics I is coupled with plane geometry. Additionally, there are certain units that are common to more than one field, such as algebraic manipulation or polynomial functions.

In this study, we considered the implementation that occurred during the March 2017–August 2017 semester in four the INACAP campuses. Additionally, for the engineering degrees, we considered the versions of Mathematics I that were shared as common units. The common units on which this study was conducted were algebraic manipulation (Alg), equations and systems of equations (Eqn. and Sys.), polynomial functions (Polyn.), and exponential and logarithmic functions (Exp/Log. Fun.).

In summary, this study focused on 229 courses taught by 59 professors covering 5,507 students. These students were from four campuses who chose Mathematics I in their engineering degrees (Table 1).

4.3 | Context of application and characteristics of the intervention

The type of feedback was randomly assigned to the 229 courses. Thus, close to half (49.8%) of these courses worked with deferred feedback (114 courses), whereas the other half (50.2%) worked with immediate feedback (115 courses) (Table 1).

TABLE 1 Population

	N. of students		N. of courses	
Venues	**Immediate**	**Deferred**	**Immediate**	**Deferred**
Curicó	391	297	14	13
La Serena	526	500	23	23
Renca	960	957	40	42
Santiago Sur	931	945	38	36
Grand total	2808	2699	115	114

The decisions made for the implementation were common to all courses. During the first three units, two quizzes were prepared—hereinafter referred to as Q1 and Q2—and only one quiz was available for the last unit: Exponential and logarithmic functions. Each quiz includes six questions that remained open for three weeks (time assigned to the units) per program of the subject.

Given the randomness in the statements, each quiz could be answered by students as many times as desired until the questionnaire was open and the grading mode selected by the system was chosen based on the highest grade obtained by a student among all the attempts.

The questionnaires were conducted outside the classroom using the institution's computer equipment or personal equipment.

4.4 | Data collection and analysis

The data used to perform the analysis were obtained from the Moodle platform, which has a record of all the attempts made by the students for answering the quiz. In particular, the entry date records, total time used, GPA, and score for each question constituted the data required for performing the analysis. These data enabled a comparison with the global grades per student as well as with the official list of enrollees in each course.

Initially, the normality in the distribution of the data was evaluated through a Kolmogorv–Smirnov test for the two feedback groups with respect to each variable.

The variable "participation rate" was based on the value of statistical significance, which is lower than 0.05 in all cases, except for the Q2 control (polynomial functions). Then, it was concluded that the data did not present a normal distribution in the rest of the variables for quizzes 1 and 2 of algebraic manipulation, equations and systems of equations, polynomial functions, and quiz 1 of exponential and logarithmic functions.

The variables "number of attempts," "working time in the platform," "GPA," and "maximum grades" were based on the value of statistical significance, which was lower than 0.05 in all cases. Then, it was concluded that the data did not present a normal distribution for quizzes 1 and 2 of algebraic manipulation, quizzes 1 and 2 of equations and systems of equations, quizzes 1 and 2 of polynomial functions, quizzes 2 of equations and systems of equations, and quiz 1 of exponential and logarithmic functions.

Considering the result of the normality test described above, we decided to use nonparametric tests. Given the nature of the variables: independent, dichotomous, and interval-dependent, we decided to use the Mann–Whitney U test, a non-parametric test of median contrast, which is equivalent to the comparison of two independent averages through the Student's *t* test, and the calculations have been performed using the SPSS V statistical program.

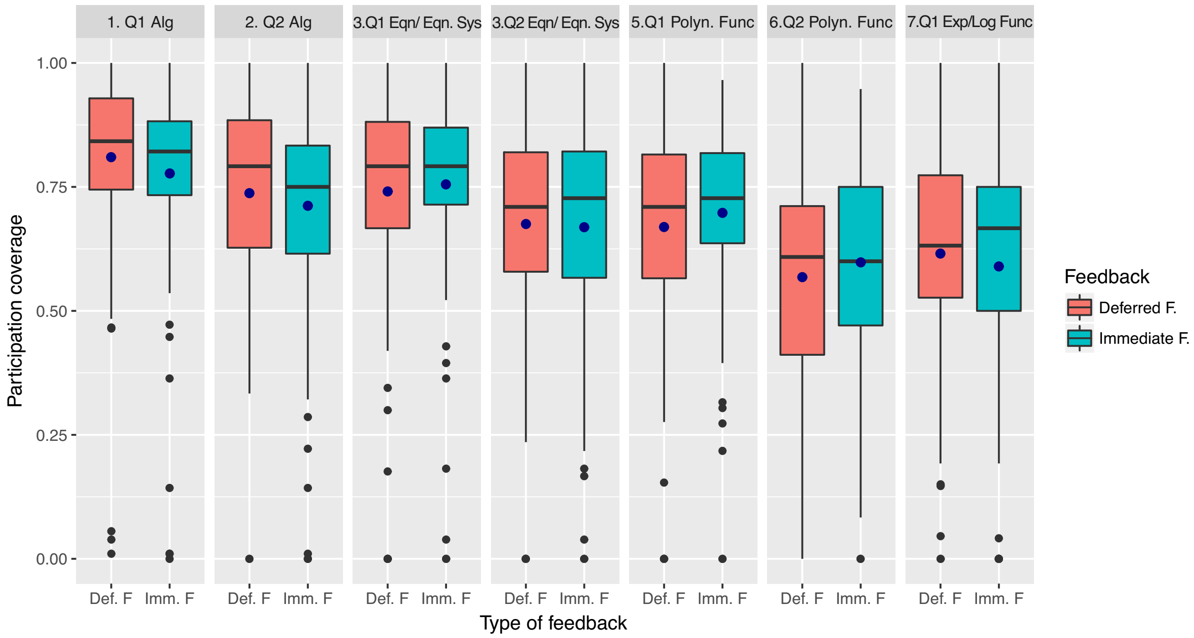

FIGURE 3 Student participation rate in SEDOL-M

For the analysis of data and presentation of the results, the following factors are combined: the type of feedback (immediate or deferred), the time of evaluation as Q1 and Q2, and the topic within the subject.

Throughout the following section, the results are presented according to the following abbreviations:

- Q1 Alg. M.: Quiz 1 of algebraic manipulation
- Q2 Alg. M.: Quiz 2 of algebraic manipulation
- Q1 Eqn. and Syst.: Quiz 1 of equations and systems of equations
- Q2 Eqn. and Syst.: Quiz 2 of equations and systems of equations
- Q1 Polyn. Func.: Quiz 1 of polynomial functions
- Q2 Polyn. Func.: Quiz 2 of polynomial functions
- Q1 Exp/Log. Func.: Quiz 1 of exponential and logarithmic functions

5 | RESULTS AND DISCUSSION

This section presents and discusses the main results linked to the activity of students outside the classroom through WIRIS-QUIZZES. Specifically, the level of participation and the number of the questionnaire attempts, the time of work in the platform, the average of grades and the highest grades were analyzed depending on the type of feedback.

5.1 | Participation does not change according to the type of feedback

One of the rules of operation was that the students could make all the attempts that they considered convenient within the term established by the professors, that is, the time during which the subject was taught in general. For calculating

TABLE 2 Comparison test and descriptive statistics of the participation rate of students on the platform

	R. deferred		R. immediate		
Quiz	**N. of courses count**	**Average participation rate (Sd)**	**N. of courses count**	**Average participation rate (Sd)**	**Sig. asymptotic (bilateral)**
Q1 M. Alg.:	111	81.02% (17.95%)	113	77.72% (18.93%)	0.082
Q2 M. Alg.:	111	73.76% (19.3%)	113	71.19% (19.57%)	0.189
Q1 Eqn. and Syst.	111	74.06% (20.83%)	113	75.51% (18.59%)	0.778
Q2 Eqn. and Syst.	111	67.52% (21.28%)	113	66.86% (20.3%)	0.721
Q1 Polyn. Func.	111	66.89% (21.16%)	113	69.75% (17.81%)	0.411
Q2 Polyn. Func.	111	56.83% (22.03%)	113	59.78% (19.97%)	0.351
Q1 Exp/Log Func.	111	61.58% (20.61%)	113	58.95% (23.17%)	0.626

*Sig <0.05; H_0 is rejected.

"participation," the number of students who made at least one attempt in an quiz was divided by the number of students enrolled in the course. Figure 3 shows the distribution of all the courses for each of the quizzes based on the type of feedback.

We concluded that there is no difference between the coverage medians of Q2 control (Polyn. Func.), a fact that corresponds to a result summarized in Table 2, which summarizes the results of the implementation of the Mann–Whitney U test.

It is worth noting that the feedback does not produce major differences in coverage: in some evaluations, the participation is slightly higher with immediate feedback than it is with deferred feedback (as in Q1 and Q2 algebra), whereas in others, exactly opposite results are obtained (as in Q1 and Q2 Polyn. Func.). However, if we observe the evolution of participation over time, in general, a downward trend is observed with the progression of the semester. We assume that this is a consequence of attrition by students who continuously work in the successive assessments.

We can also appreciate that in the first assessments of each unit there is a slight rebound with respect to the assessment 2 of the previous unit. For example, participation in the first assessment of Eqn. and Syst. is greater than in the second assessment of Alg. M., both with immediate feedback and deferred feedback. This could be explained because, according to the report by each location, they decided to open the two quizzes of each unit during the whole period of classes during which the respective unit was worked, that is, Q1 and Q2 of each unit opened and closed on the same day; this could have caused a greater focus on Q1 in each unit.

5.2 | Further attempts with immediate feedback

As each assessment was open for a period in which students could answer as many times as they wanted, the number of attempts the students made in each assessment was analyzed and differentiated according to the type of feedback (see Figure 4).

As in the results of the participation rate, there are a greater number of attempts in Q1 than in Q2 in each of the subjects. In this case, there is no downward trend as the semester progresses; rather, the number of attempts varies (between 3 and 5) depending on the topic being worked with, and this may be related to the difficulty of each specific subject for the students. For example, for both immediate and deferred feedback, the maximum average of attempts was higher in the first quiz of polynomial functions.

When applying the Mann–Whitney U comparison test, we appreciate that there are significant differences between immediate and deferred feedback, in terms of the number of attempts made by the students; this is true in all cases except Q1 in the algebra unit. If we look at Table 3, we see that the difference is in favor of immediate feedback.

Regarding the distribution of the students per the number of attempts, it is practically identical in the first three assessmentss. From the fourth quiz, a much greater increase in the number of attempts of the students whose assessments were with immediate feedback is observed.

This increase in the number of attempts could be explained by what was described by the teachers whose students worked with immediate feedback. As reported by them, the students answered each of the questions of a particular assessment, and if in some of them the system

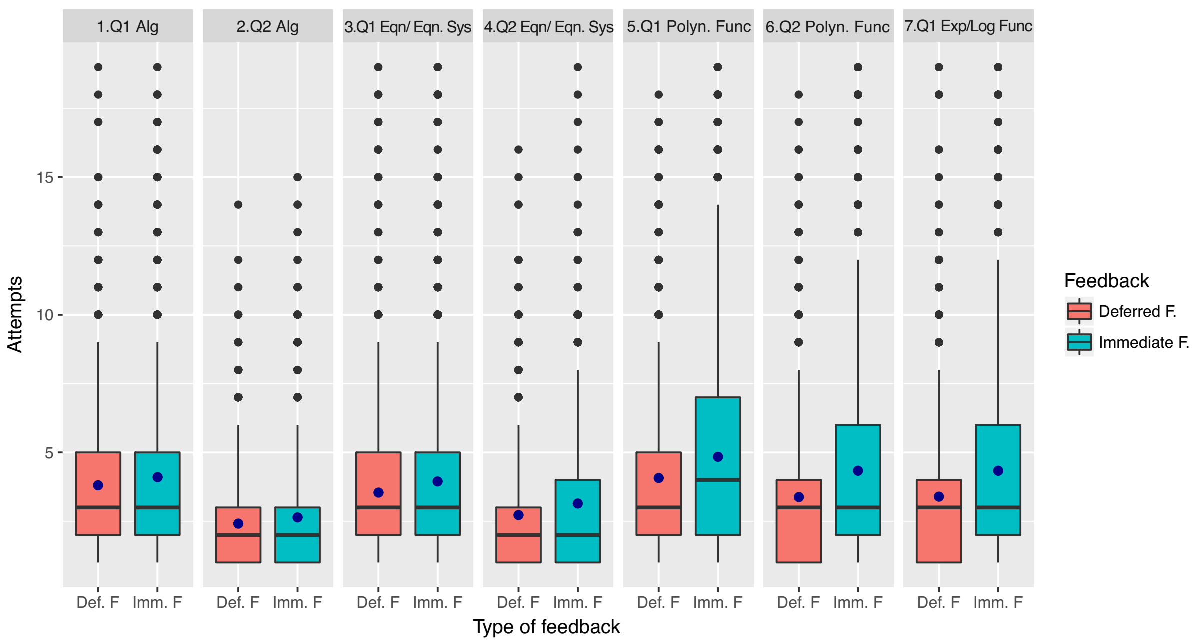

FIGURE 4 Number of attempts in each evaluation

TABLE 3 Comparison test and descriptive statistics of the number of attempts per student on the platform

Quiz	R. deferred N. of students count	R. deferred N. of attemps mean (Sd)	R. immediate N. of students count	R. immediate N. of attemps mean (Sd)	Sig. asymptotic (bilateral)
Q1 M. Alg.:	2165	3.81 2.57	2251	4.1 3.06	0.102
Q2 M. Alg.:	1952	2.41 1.6	2044	2.64 1.94	0.019
Q1 Eqn. and Syst.	1952	3.54 2.84	2128	3.94 3.28	0.000
Q2 Eqn. and Syst.	1785	2.73 1.95	1900	3.14 2.41	0.000
Q1 Polyn. Func.	1770	4.07 (3)	1940	4.84 (3.83)	0.000
Q2 Polyn. Func.	1510	3.38 (2.69)	1673	4.33 (3.71)	0.000
Q1 Exp/Log Func.	1619	3.39 (2.65)	1655	4.34 (3.55)	0.000

*Sig <0.05; H_0 is rejected.

indicated that their answer was incorrect, the students finished the attempt and started another from scratch, since the "cost" of this attempt—in terms of a grade—was null. This, in turn, caused an increase in the number of attempts concentrated in the first half of the questions. On the other hand, in the case of delayed feedback, the student had to enter all the answers, in order to know which of them were correct.

5.3 | Longer work time with immediate feedback

Figure 5 summarizes the distribution of work time on the platform, calculated as the number of minutes that each student used to answer all the attempts made in each of the assessments, according to the type of feedback. The graph that shows all the students who worked up to 500 minutes corresponding to 99.9% of the population was chosen since there were students who worked up to 1,000 min. However, including them in the graph did not allow the distribution of the majority of students to be clearly observed.

In the graph we can observe that the amount of time of work in the platform is greater in the assessments with immediate feedback than in the assessments with delayed feedback. This tendency is also seen in the values of means of Table 4. When applying the Mann–Whitney U comparison test, it is confirmed that these differences between immediate and deferred feedback, in terms of the amount of work time on the platform, are statistically significant.

This difference can be associated with the variable "number of attempts" since a greater number of attempts implies a longer working time on the platform. However, unlike the variable "attempts," the distribution of quartiles is different in all quizzes and not only from the fourth quiz, as in the case of the number of attempts.

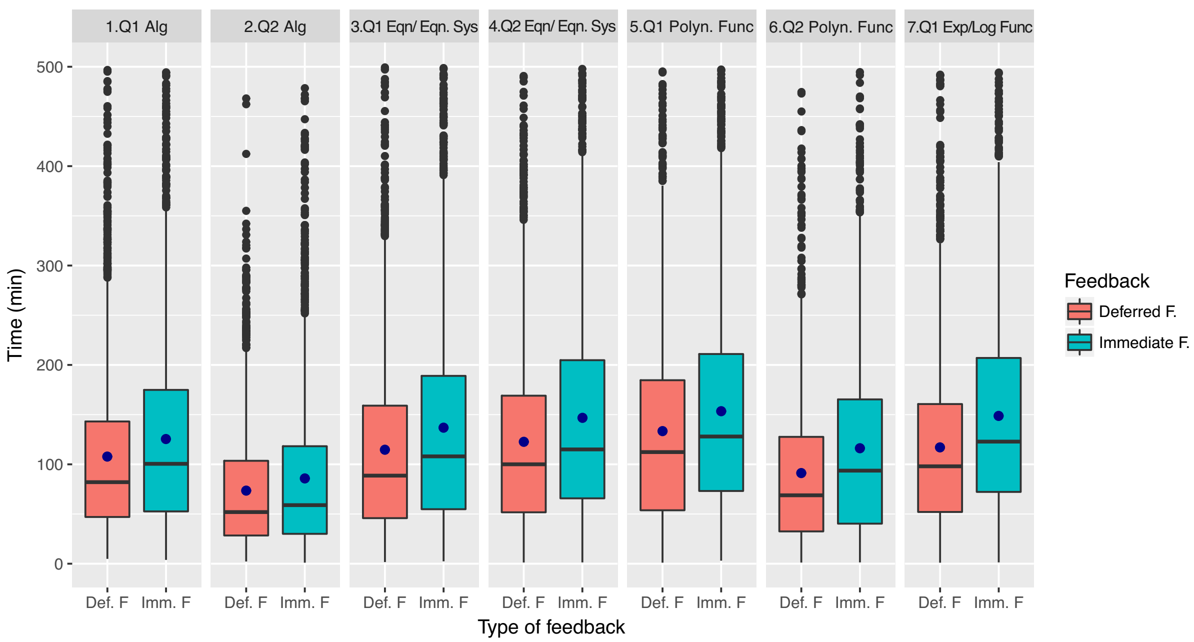

FIGURE 5 Work time on the platform

TABLE 4 Comparison test and descriptive statistics of the work time per student on the platform

Quiz	R. deferred		R. immediate		Sig. asymptotic (bilateral)
	N. of student count	Minutes mean (Sd)	N. of students count	Minutes mean (Sd)	
Q1 M. Alg.:	2167	109.11 (88.73)	2257	130.46 (107.47)	0.000
Q2 M. Alg.:	1952	73.88 (64.11)	2044	87.02 (82.02)	0.000
Q1 Eqn. and Syst.	1955	120.56 (107.16)	2150	152.17 (142.33)	0.000
Q2 Eqn. and Syst.	1785	126.77 (103.49)	1903	158.77 (134.83)	0.000
Q1 Polyn. Func.	1782	139.99 (114.81)	1979	172.72 (146.27)	0.000
Q2 Polyn. Func.	1514	92.43 (82.24)	1694	122.82 (112.76)	0.000
Q1 Exp/Log Func.	1623	121.29 (101.31)	1672	160.44 (129.74)	0.000

* Sig <0.05; H_0 is rejected.

5.4 | The GPA of all attempts in each assessment is lower with immediate feedback

Let us recall that the students had unlimited attempts on the platform, and for each of them they received a grade. From all the attempts that a student made in an quiz, an average is obtained. Figure 6 shows the distribution of the averages of all the attempts made by the students in each quiz, according to the type of feedback. The rating on the platform varies from 0 to 100.

The graph shows that the GPA in the assessment with immediate feedback is lower than the average of the assessment grades with delayed feedback. This is also shown in Table 5, which displays the means of each assessment, according to the type of feedback.

When applying the Mann–Whitney U comparison test (last column, Table 5), it is confirmed that the differences between the immediate and deferred feedback, as regards the GPA, are statistically significant.

This could be explained by the same dynamics of immediate feedback, since the student had the correction information for each answer before the end of the exercise. This, coupled with the null cost of this attempt in terms of a grade, may have encouraged the student to restart the assessment from scratch.

On the other hand, in assessments with delayed feedback, the student had to answer all the questions before knowing which ones were correct or not; this produced a smaller number of attempts, although these attempts were more efficient in terms of a grade.

The graph also shows a general downward trend in grades over the period, although in the assessment 1 of each unit the decrease is attenuated. This tendency could be explained by the increasing difficulty of the units.

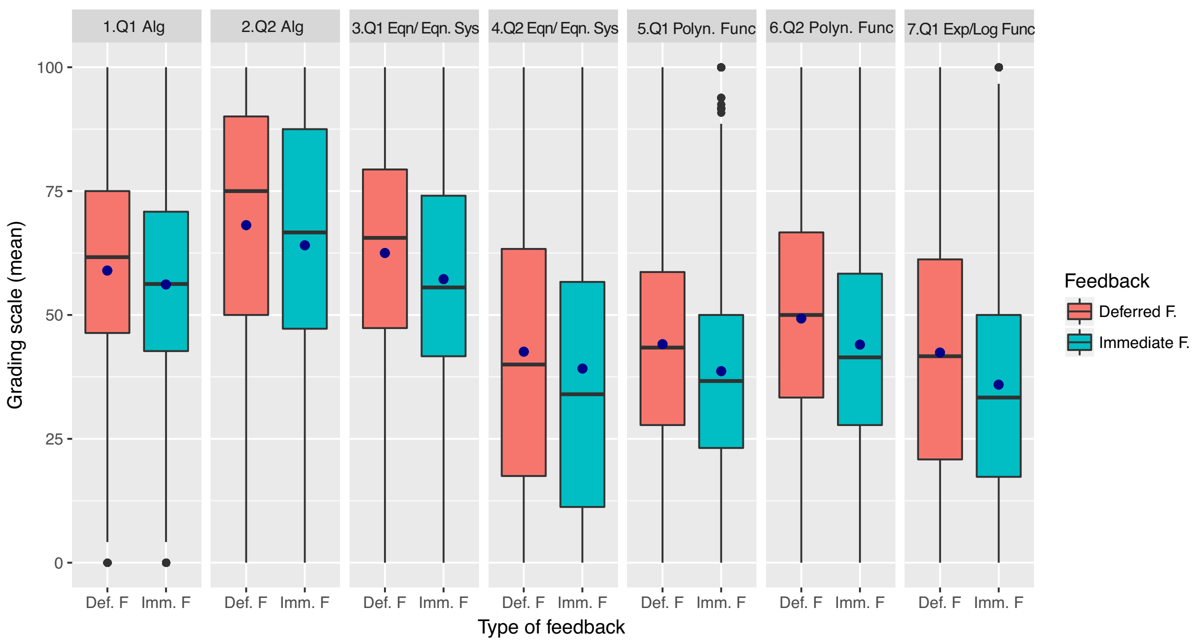

FIGURE 6 Average of all grades

TABLE 5 Comparison test and descriptive statistics of the GPA of all attempts per student on the platform

Quiz	R. deferred N. of students count	R. deferred Average rating mean (Sd)	R. immediate N. of students count	R. immediate Average Rating mean (Sd)	Sig. asymptotic (bilateral)
Q1 M. Alg.:	2165	59 (22.9)	2251	56.2 (22.2)	0.000
Q2 M. Alg.:	1952	68.1 (26.7)	2044	64.1 (27.4)	0.000
Q1 Eqn. and Syst.	1952	62.5 (23.9)	2128	57.4 (23.8)	0.000
Q2 Eqn. and Syst.	1785	42.6 (31.4)	1900	39.2 (31.2)	0.000
Q1 Polyn. Func.	1770	44.2 (24.7)	1940	38.9 (23.6)	0.000
Q2 Polyn. Func.	1510	49.3 (26.4)	1673	44.2 (25.3)	0.000
Q1 Exp/Log Func.	1619	42.4 (28.3)	1655	36.1 (25.5)	0.000

*Sig <0.05; H_0 is rejected.

5.5 | The maximum grade of the attempts in each assessment is slightly higher with immediate feedback

As explained above, the students had an unlimited number of attempts while each assessment was open, and the grade is calculated as the highest grade among all the attempts made during the period. For this reason, this indicator is different from the average of all the ratings, which was analyzed in the previous sub-section. Figure 7 shows the distribution of the maximum grades for each of the assessments and according to the type of feedback.

In all the assessments, the maximum score is slightly higher in the assessments, with immediate feedback than with deferred feedback, between 0.6 and 3.7 percentage points (see Table 6). We can also see that the distribution of the notes in quartiles is practically identical in almost all assessments, except for the first and the last. We can also show that the grades are better in the first assessments than in the last ones, and in turn, there is a greater dispersion in the students' grades as of the fourth assessment.

In general, there is a downward trend in the maximum grades as the semester advances; however, this decrease is less pronounced than in the case of the average of all assessments.

These data confirm the difficulty faced by students of working with the contents that are presented as the semester advances. As we saw that in the last subjects, there is an increase in the number of attempts, but a decrease in the final grade. This increase in difficulty may partly explain the decrease in student participation as time passes.

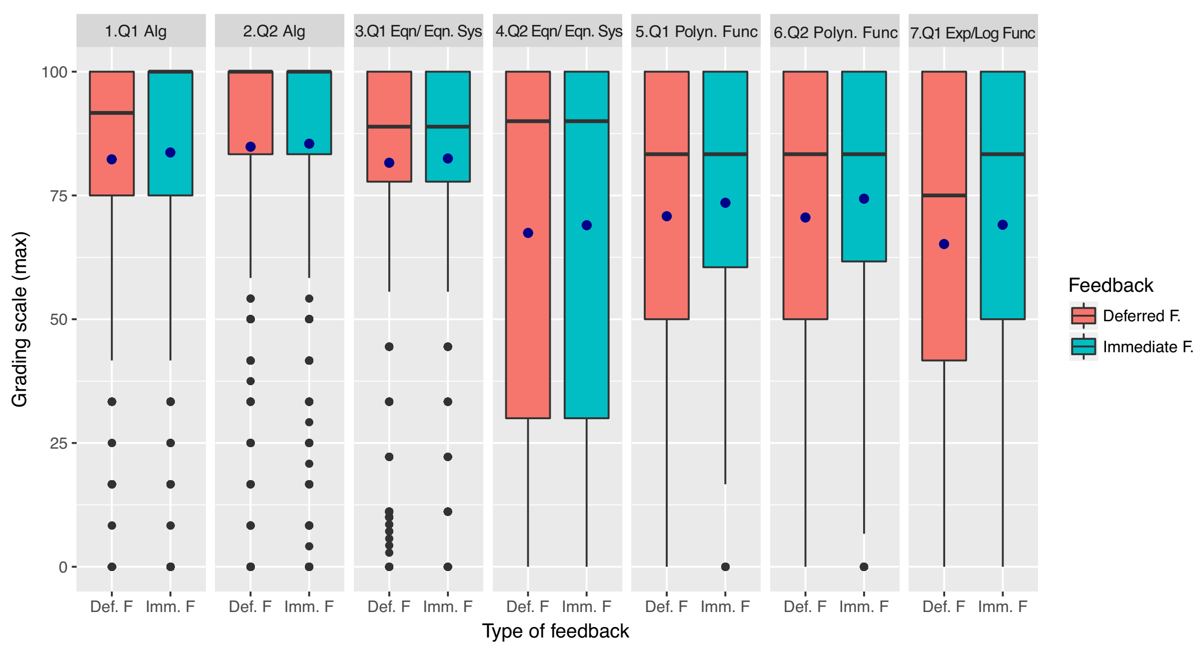

FIGURE 7 Distribution of maximum grades in each evaluation and according to the type of feedback

TABLE 6 Comparison test and descriptive statistics of maximum grades of all attempts per student on the platform

	R. deferred		R. immediate		
Quiz	N. of students count	Maximum grade mean (Sd)	N. of students count	Maximum grade mean (Sd)	Sig. asymptotic (bilateral)
Q1 M. Alg.:	2165	82.3 (25.8)	2251	83.7 (24.9)	0.024
Q2 M. Alg.:	1952	84.8 (25.9)	2044	85.4 (26.6)	0.018
Q1 Eqn. and Syst.	1952	81.5 (24.4)	2128	82.3 (24.9)	0.011
Q2 Eqn. and Syst.	1785	67.4 (38.3)	1900	69 (39.3)	0.004
Q1 Polyn. Func.	1770	70.6 (30.4)	1940	73.1 (31.2)	0.000
Q2 Polyn. Func.	1510	70.5 (30.9)	1673	74.1 (29.8)	0.000
Q1 Exp/Log Func.	1619	65.1 (35.2)	1655	68.8 (35.9)	0.000

*Sig <0.05; H_0 is rejected.

When applying the Mann–Whitney U comparison test, it is confirmed that these differences between immediate and deferred feedback, as to the average of the maximum score obtained among all the students' attempts, are statistically significant. These results are presented in the last column of Table 6.

6 | CONCLUSIONS AND NEXT STEPS

This research is part of a project that seeks to make an online assessment system available to students. For engineering students, with regard to a mathematics-leveling course, this system serves as an autonomous work system outside the classroom. The two basic pillars supporting this project consider that learning occurs when these students actually perform the mathematical work [7] and when there is a feedback that is delivered in a relevant and timely manner [13].

Feedback can vary in its nature and at the time it is delivered [13]. In this project, the students had to answer questions designed with random parameters in the statement and the system provides two types of feedback: task, which indicated to them if their answer was correct or not and process, which gives them the step-by-step development of the question according to the random parameters of the statement.

In this research, the effect that the moment in which feedback is delivered on students' work is studied. About half of the students had access to assessments with immediate feedback—of task and process—this feedback could be accessed by a button on each question. The other half of the students had access to assessments that were configured with deferred feedback, that is, once the students answered all the questions, the system gave them the task and process feedback.

The study took place in 7 assessments applied during a semester for 4 different topics: algebra, equations and systems of linear equations, polynomial function and exponential and logarithmic function. Each assessment contained 6 questions.

The work of the students was scrutinized using five variables: the participation rate, the number of attempts (the students could answer the assessment as many times as they wanted during the time it was open on the platform), the work time, the GPA and the maximum grade obtained in each assessment.

The results show that the participation rate throughout the semester does not change according to the moment in which the feedback is delivered, and for both cases it is observed that this rate is high (on average 68.67%). However, there is a decrease in participation as the semester advances; we attribute this to a consequence of fatigue undergone by the students because they are subject to a demanding and sustained working rhythm throughout their interaction with the system.

When participation is analyzed, it was observed that in the case of assessments with immediate feedback, the number of attempts (on average 0.6 more attempts per assessment), the work time (on average 28 more minutes of work per assessment) and the maximum rating (on average 2% more per assessment) are higher than for the case of deferred feedback. The only indicator that is higher for the case of deferred feedback is the GPA. The explanation for this is that in the assessments with immediate feedback, the student—upon receiving information about each question—restarted the questionnaire when he made an error, since the attempts had no cost.

The foregoing shows that when students receive immediate feedback, they choose a strategy that implies a higher cost in terms of number of attempts and work time and whose benefit is a small increase in the maximum grade. Another related cost is that despite increasing the practice, the effort is concentrated on the first questions—unlike the assessment with deferred feedback, where students spend less time and attempts, get a slightly lower grade, but their practice is distributed throughout the assessment instead of the first questions only. If we consider that the interest of the project is to generate a system of study and not only an assessment system, delayed feedback seems more appropriate for that purpose.

This study shows that when implementing a technology that has the provision for changing the teaching–learning dynamics in a simple manner, particularly when automatic feedback is incorporated, a change can be observed in the behaviors of the students, based on which they develop different strategies to supervise themselves; this is consistent with the main objectives of feedback processes [18].

It should be noted that the feedback provided by the system is framed in a "traditional" conception to some degree, since it comprises providing information, from which the student is expected to individually reflect and improve his/her performance; this could be regarded as a concept of feedback surpassed by more participatory models [4]. However, once the usefulness of the system is noticed, its most significant virtues are its sustainability and scalability—feedback covers a large volume of students, something impossible to conduct with traditional teaching methods. In particular, the information is specific to each student, which allows adjusting according to their performance, immediacy-feedback is timely, that is, the student does not have to wait for it, as he/she receives it automatically. Given the importance of evaluation in the teaching–learning process, having an automatic feedback system is an important first step to continue in the direction of improvement of the institution.

This first step is only the starting point of a line of research aimed at analyzing the effectiveness of the use of this type of questionnaire for learning mathematics in technical studies. Fundamentally, the purpose is to verify whether the student's performance relies on the type and quality of the homework being proposed and the feedback they receive or on the teacher's experience and student's skills [19]. In particular, exploring the extent to which automatic feedback encourages positive motivational beliefs and self-esteem among students in this particular case is a considerably interesting topic of research.

ACKNOWLEDGMENT

This article is a part of the research and innovation project SEDOL-M, which is funded by the CIEDU Education Innovation Center of INACAP, Chile.

ORCID

Jorge Gaona http://orcid.org/0000-0001-6367-529X
Mercedes Reguant http://orcid.org/0000-0002-0830-7854

REFERENCES

1. M. Abboud-Blanchard, Teachers and technologies: Shared Constraints, Common Responses, *The mathematics teacher in the digital era*, (A. Clark-Wilson, O. Robutti, and N. Sinclair eds.), Springer, Netherlands, 2014, pp. 297–317.
2. M. Artigue, The future of teaching and learning mathematics with digital technologies, *Mathematics education and technology-rethinking the terrain*, (C. Hoyles and J.-B. Lagrange ed.), Springer, Boston, MA, 2010, pp. 463–475.
3. S. Bonham, D. Deardorff, and R. Beichner, *Comparison of student performance using web and paper-based homework in college-level physics*, J. Res. Sci. Teach. **40** (2003), 1050–1071.
4. D. Boud and E. Molloy (eds.), Decision-making for feedback, *Feedback in higher and professional education: Understanding it and doing it well*, Routledge, London, 2013, pp. 202–217.
5. E. Cano, *Análisis de las investigaciones sobre feedback: aportes para su mejora en el marco del EEES*. Bordón Revista de Pedagogía **66** (2014), 9–24.
6. C. Cazes, G. Gueudet, and F. Vandebrouck, *Using E-exercise bases in mathematics: Case studies at university*, Int. J. Comput. Math. Learn. **11** (2006), 327–350.
7. Y. Chevallard, J. Gascon, and M. Bosh, *Estudiar matemáticas el eslabón perdido entre enseñanza y aprendizaje*, I.C.E., Universitat de Barcelona, 2006.
8. F. Díaz-Barriga and F. Arceo, *Enseñanza situada: vínculo entre la escuela y la vida*. Ciudad de México: McGraw Hill, 2006.
9. T. Escudero, *Desde los tests hasta la investigación evaluativa actual. Un siglo, el XX, de intenso desarrollo de la evaluación en educación*, Revista Electrónica de Investigación y Evaluación Educativa, **9** (2003), 11–43.
10. A. Fernández, *La evaluación orientada al aprendizaje en un modelo de formación por competencias en la educación universitaria*, Revista de Docencia Universitaria, **8** (2010), 11–34.
11. J. Gaona and C. Hardy, La evaluación dinámica como motor de aprendizaje incorporando Wiris en Moodle. Un ejemplo de nivelación en matemáticas, (J. Asenjo, O. Macías, and J. Toscano eds.), Congreso Iberoamericano de Ciencia, Tecnología, Innovación y Educación, OEI, Buenos Aires, 2014.
12. E. Guba and Y. Lincoln, *Fourth generation evaluation*. Sage, Newbury Park, 1989.
13. J. Hattie and H. Timperley, *The power of feedback*, Rev. Educ. Res. **77** (2007), 81–112.
14. S. Jang, *Exploration of secondary students' creativity by integrating web-based technology into an innovative science curriculum*, Comput. Educ. **52** (2009), 247–255.
15. J. Johnson and R. Mckenzie, *The effect on student performance of web- based learning and homework*, J. Economic. Economic Educ. Res. **14** (2012), 115–126.
16. I. Jones, *Computer-aided assessment questions in engineering mathematics using MapleTA®*, Int. J. Math. Educ. Sci. Technol. **39** (2008), 341–356.
17. K. Leong, *College students attitude and mathematics achievement using web-based homework*, Eurasia J Math. Sci. Technol. Educ. **10** (2014), 609–615.
18. D. Nicol and D. MacFarlane-Dick, *Formative assessment and selfregulated learning: A model and seven principles of good feedback practice*, Stud. Higher Educ. **31** (2006), 199–218.
19. S. Palocsay and P. Stevens, *A study of the effectiveness of web-based homework in teaching undergraduate business statistics*, Decis. Sci. J. Innov. Educ. **6** (2008), 213–232.
20. J. Roschelle et al., *Online mathematics homework increases student achievement*, AERA Open **2** (2016), 1–12.

21. C. Sangwin et al., Micro-level automatic assessment supported by digital technologies, *Mathematics education and technology-rethinking the terrain*, (C. Hoyles and J.-B. Lagrange eds.), Springer US, 2010, pp. 227–250.
22. K. Stacey and D. Wiliam, Technology and assessment in mathematics, *Third international handbook of mathematics education*, Springer, New York, 2013, pp. 721–751.
23. T. Sancho-Vinuesa, N. Escudero-Viladoms, and R. Masià, *Continuous activity with immediate feedback: A good strategy to guarantee student engagement with the course*, Open Learning **28** (2013), 51–66.
24. UTCI, Informe de Autoevaluación Institucional., Santiago de Chile, 2016.
25. M. Yorke, *Formative Assessment in Higher Education: Moves Towards Theory and the Enhancement of Pedagogic Practice Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice*, Higher Educ. **45** (2014), 477–501.

J. Gaona is a PHD student at Didactiques de Mathématiques program at Université Paris Diderot, France, he has done Master Recherche in Didactiques des Mathématiques at Université Paris Diderot (2015), and is graduate in Mathematics from the Pontificia Universidad Católica de Valparaíso, Chile (2008). He is currently a researcher at LDAR Laboratory in France and consultant at CIEDU in Universidad Tecnológica de Chile INACAP. He won a scholarship at CONICYT from the Chilean government.

M. Reguant is a PhD in Education. She is an associate professor of Research Methods and Diagnosis in Education at the University of Barcelona. SAR-ICE consultant (Research Consultancy Service). She is also an RIMDA-UB consultant (Research, Innovation and Improvement of Teaching and Learning independent methodological consultant, and member of the Teaching Innovation Group MideMe.

I. Valdivia received his master in Statistics from the Pontificia Universidad Católica de Valpaíso, Chile and master in Territorial Environmental Management and Forestry Engineer from the Universidad de Talca, Chile, with high experience and specialization in the area of quantitative methods applied to the field of research and analysis of data. His academic development is associated with the teaching of mathematics and statistics for more than 10 years at Universidad Tecnológica de Chile INACAP, teaching and developing research with teachers and students, and also collaborating with other universities.

M. Vásquez is a PhD in Education. He is professor of Research Methods in Education at the Technological University of Chile INACAP and a coordinator of the On Line Mathematics Evaluation System Project (SEDOL-M).

T. Sancho-Vinuesa is senior lecturer of Mathematics at Universitat Oberta de Catalunya (UOC) in Barcelona, Spain, where she teaches Mathematics for Engineers and conducts research on e-assessment, feedback and learning analytics as head of the LAIKA (Learning Analytics for Innovation and Knowledge Application in Higher Education) Group.

How to cite this article: Gaona J, Reguant M, Valdivia I, Vásquez M, Sancho-Vinuesa T. Feedback by automatic assessment systems used in mathematics homework in the engineering field. *Comput Appl Eng Educ*. 2018;1–14. https://doi.org/10.1002/cae.21950