International Journal of Engineering Education Vol. 39, No. 4, pp. 949–960, 2023
Printed in Great Britain

0949-149X/91 $3.00+0.00

Automated Intelligent Feedback Based Learning in a Software Development Project Management Course*

JUAN J. LÓPEZ-JIMÉNEZ, JOSÉ A. GARCÍA-BERNÁ, AMBROSIO TOVAL and JOSÉ L. FERNÁNDEZ-ALEMÁN
Faculty of Computer Science, Department of Informatics and System, University of Murcia, Campus of Espinardo, Murcia, Spain.
E-mail: juanjoselopez@um.es, josealberto.garcia1@um.es, atoval@um.es, aleman@um.es

ALI IDRI
Software Project Management Research Team, ENSIAS, Mohammed V University in Rabat, Rabat, Morocco.
E-mail: idri.ali123@gmail.com

This paper presents an empirical study of a formative assessment approach based on intelligent diagnostic feedback. A public audience response system called SIDRA was integrated with clustering-based data analysis to generate diagnostic feedback for guided learning. A total of 138 computer science students enrolled in a Software Development Project Management course during the 2021/2022 academic year were taught using two different strategies. Eighty students in the experimental group used intelligent SIDRA (i-SIDRA), while 58 students in the control group received the same training but without using i-SIDRA. A statistically significant difference in final exam grades was found between students using i-SIDRA versus a traditional teaching methodology ($U = 3306$, $p < 0.001$). No statistically significant differences were obtained in the final grades between the elaborated feedback group and the reduced feedback group in the two experiments ($T(5,190) = –1.928$, $p = 0.110$ and $U = 443$, $p = 0.474$) conducted to evaluate the effectiveness of the two types of feedback at the end of the semester. No statistically significant differences were reached in the increase of correct questions from the first to the last attempt during the feedback-guided learning process, between the elaborated feedback group and the reduced feedback group in the two experiments carried out ($T(19) = 0.217$, $p = 0.831$ and $U = 699$, $p = 0.542$). In a questionnaire rated on a five-point Likert-type scale, students stated that i-SIDRA feedback promotes clarification and understanding of concepts ($MD = 4$).

Keywords: e-learning; feedback; intelligent systems; project management

1. Introduction

Currently, the use of new technologies in education is becoming essential in a multitude of subjects. With the arrival of the COVID-19 health crisis, it has acquired even more prominence [1, 2]. This change implies advantages such as being able to perform online teaching and the convenience of having classes and materials available at any time and place, although disadvantages such as inefficiency and difficulty in maintaining academic integrity are also found [3]. One of the resources used is the Audience Response System (ARS), with a trend in recent years to use freely available software and students' mobile devices [4]. These systems make it possible to (1) promote active learning [5–7]; (2) more knowledge is acquired than with traditional classes [8]; (3) improve motivation [9]; (4) create an environment for shy students [10, 11]; (5) monitor class attendance; (6) increase collaboration between teacher and student and even among students themselves; and (7) perform student evaluations.

ARS is a technology that allows a group of students to transmit and individually respond electronically to a questionnaire previously prepared by the teacher [12]. Numerous studies have shown that it is well received by students and teachers [4, 13, 14]. The traditional way of using these systems consisted of acquiring expensive radiofrequency equipment. In order to reduce the high costs of acquisition and maintenance of this equipment [15], several applications for mobile devices emerged to emulate these systems, such as Kahoot, Wooclap, Socrative and Arsyc [16–19]. Among them we find the SIDRA (Sistema de Respuesta Inmediata de la Audiencia) tool that was one of the first applications of this type to appear (https://docentis.inf.um.es/sidra) available since 2011, and which has been successfully used in university teaching [20, 21].

ARS can integrate artificial intelligence through neural networks or clustering systems that allow incorporating cognitive diagnostic techniques to improve learning [22–25]. The combination of both resources provides feedback to the student for instant learning and assessment. SIDRA has a module called i-SIDRA that allows grouping students into knowledge clusters according to the answers to the questions of a multiple-choice question (MCQ) test, to return personalized intelligent diagnostic feedback to the student.

This paper mainly presents two contributions: (1) an experiment comparing the academic performance achieved by computer science students using a multiplatform intelligent feedback system (i-

* Accepted 7 April 2023.

SIDRA), with that obtained using a traditional approach; (2) an analysis of the effect of the type of feedback (reduced or elaborated) on the students' performance in a MCQ based intelligent feedback system. To the best of our knowledge, this is the first work presenting an empirical study with these learning systems in a higher education Computer Science context.

2. Feedback in Teaching

Feedback is a key component of building scaffolds for learning. Feedback can provide information on the achievement of learning objectives and the improvement of self-regulated skills. However, delivering feedback is a difficult task for instructors, especially in crowded classroom contexts. Therefore, a number of automatic feedback systems have been proposed to alleviate the instructor's workload. In particular, feedback in computer-based learning environments is one of the essential elements for student learning. Some authors [26, 27] define feedback as the information provided by the computer about the learner's performance. Both information on the state of learning and guides to progress from the current state to the knowledge target can be included [28]. The scientific literature has been pointed out that inadequate feedback can even hinder learning [29, 30].

Three key characteristics of feedback are content, timing and presentation. These characteristics are determinants of effective feedback. Some authors [27, 31] have pointed out that the feedback factor that has the greatest influence on the learner's learning is the content. According to its complexity [27], a distinction is made between simple feedback and elaborated feedback. The former includes checking whether the answer is correct or even providing the correct answer. In the second, additional formative information can be provided in the form of hints, explanations, examples, and problem-solving strategies. However, research on the effects of both kind of feedback is scarce, so further research is required in the area. Moreover, the effectiveness of feedback can also be influenced by the format of the question [32, 33]. Some authors [34] suggest that MCQs and essay questions trigger different cognitive demands, whereby learners also show differences in perceiving and processing the information provided in the feedback [32].

In higher education, feedback should be used to shape students as self-regulated trainees [35], that is, learners should be able to regulate issues of their thinking, motivation and behaviour during learning. In the field of engineering education, few studies introducing feedback in the teaching strategy has been conducted. Most of these papers do not analyse different kind of feedback [36–39]. One of the few exceptions is the work developed by Jaeger and Adair [40] which compared the impact of reflective feedback versus corrective feedback on learning effectiveness in Mechanical and Aerospace Engineering students. This study concluded that corrective feedback promoted students to memorize the correct answers, while reflective feedback contributed to a deeper understanding of the underlying concepts. Specifically in the context of computer science education, there has been a rising interest in understanding how self-regulation skills contribute to student achievements [41]. Guidelines for good feedback practice to facilitate self-regulation has been proposed in literature [42]. To the best of our knowledge, these principles have not been applied to provide MCQ based feedback in the context of computer science education, in general, and in software development project management [43, 44], in particular. This paper aims to take a step further in the use of feedback in education and investigate the two different kinds of feeding back in a MCQ based intelligent feedback system, considering good practical principles to enable self-regulation in students. In this work, the "Assessment and feedback" strategy to foster self-regulated learning in higher education is adopted [45].

3. The i-SIDRA Learning System

The SIDRA system is a free public application with a client-server architecture. This tool allows instructors to create, collect and analyse answers to MCQs. SIDRA has interfaces for students and instructors. Users can access it via the web or mobile devices. The mobile version of SIDRA can be downloaded from the Apple App Store and Google Play [46, 47].

In SIDRA, we can provide questionnaires consisting of a set of MCQs. Students can answer the questions and see the percentage of correct answers for each question at the end or during the session. Alternatively, teachers can create quizzes, launch them for students to answer, export test results and view information about student responses. A quiz can be configured for students to participate anonymously or being identified. SIDRA has an intelligent module (i-SIDRA) that allows learners and instructors to make use of an intelligent feedback system. The following are the steps to be taken in the learning process with i-SIDRA.

3.1 Questionnaire Creation and Feedback Association

The instructor initially creates a questionnaire with several MCQs in the i-SIDRA tool. For each question, the feedback associated with each incor-

Table 1. Example of knowledge groups generated through clustering

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Group 1	a, b, c	*	*	b, c, d	b, c, d	a, b, d	*	a, b, d	*	a, b, d
Group 2	c	a, b	b	d	c, d	b	a, d	b, d	a	a, b
Group 3	c	a	b	d	c	b	d	d	a	a
Group 4	c	a	b	d	c	b	d	d	a	a
Group 5	c	a	b	d	c	b	d	d	a	a
Group 6	a	b	b	c	d	b	c	d	c	b
Group 7	*c*	*a*	*b*	*d*	*d*	*b*	*d*	*d*	*a*	*a*

rect answer should be entered. Text, images, web links, links to videos or audios can be added to the feedback.

3.2 Generation and Loading of Knowledge States

After creating the questionnaire, the knowledge groups are generated. The clusters are created based on the previous answers to the questionnaire made by other students who have already answered it. A hierarchical clustering algorithm [22, 48] creates the groups, representing knowledge states, from common answers. The clusters contain the same correct and/or incorrect answers to the questions posed, and constitute the participants' response patterns. The number of clusters is calculated using the Silhouette method [49]. The tool i-SIDRA shows the instructor graphically the possible clusters by means of a dendrogram. The teacher can select the number of groups considered appropriate for the test or accept the number of groups suggested by the tool following the Silhouette method. Table 1 shows an example for a 10-question test with 4 possible answers ('a', 'b', 'c' and 'd') in which 7 groups have been formed. The character '*' stands for any of the four answers. For example, group 1 represents students who answered a, b or c in question P1, any answer in questions P2 and P3, and so on. The system automatically generates feedback for each group according to the deficiencies represented in that group. This feedback is composed of tips, references, hints and guidance material that a student classified in a group (or knowledge state) needs to read in order to understand certain concepts based on their given answers. The automatically generated feedback for each group can be modified by the teacher to avoid duplication or for other purposes. After this step, the teacher can schedule a test for a specific date and duration.

3.3 Student Learning

Students will be able to access and answer the questionnaire during the time period set for completion. The submission of answers to the questionnaire is done jointly for all the questions in the questionnaire. With each submission of answers, the student's knowledge group or state is recalculated. If the learner's answers are not part of the group with the highest knowledge level (or perfect knowledge state), the application will show the learner the feedback for that group, thus inviting learner to answer again. The learner will then have to read and analyse the feedback to correct them answers and resubmit the answers as many times as required until the perfect knowledge state is reached or until the time available for the test expires (see Fig. 1).

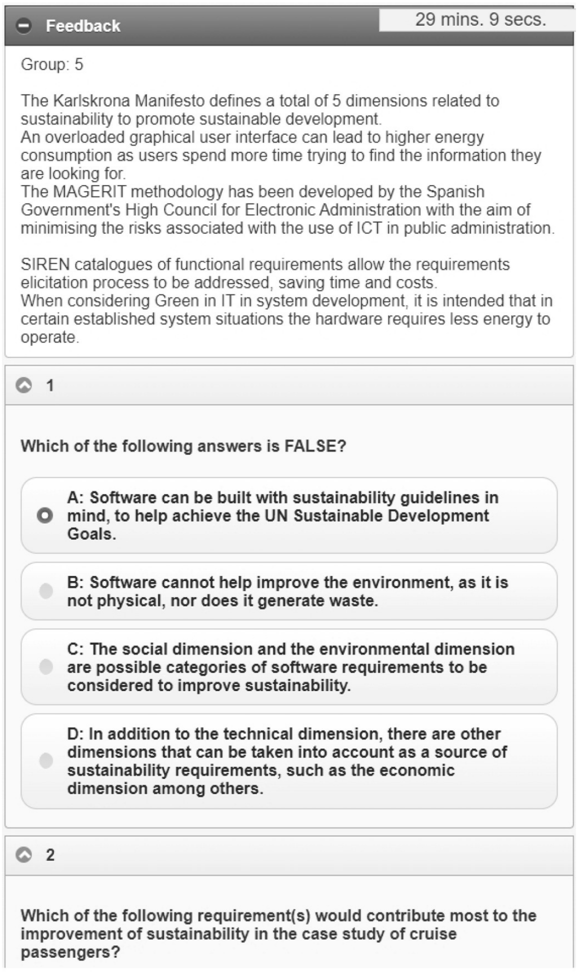

Fig. 1. Example of questions and feedback in the student interface.

3.4 Analysis of the Results by the Teacher

When students' learning phase is over, the teacher will be able to check all the information of the test. In particular, it will be possible to:

- View in list form the students who have accessed the questionnaire and the last access date.
- View in list form the students who have reached the perfect knowledge state and the last submission date.
- View each one of the submissions during the test for each student, being able to see at what time the test was ran, the answers the students sent and the state reached with each submission.
- View in the form of a state diagram, individually and collectively, the evolution of the students with data and statistics, and the feedback received by a student in each knowledge state (see Fig. 2).
- Reproduce the states of knowledge in which each student has been located. The teacher can click on the Play button to automatically highlight the current knowledge state and related information after each submission (see Fig. 3).

3.5 How the System Learns

The questionnaire can be launched to as many groups of students as necessary without generating new clustering groups again. However, with each test, new answers are obtained, which provides feedback to the system itself and a new clustering can be generated.

4. Method

4.1 Participants and Data Collection

This prospective non-randomized controlled trial involved a convenience sample of students from the University of Murcia enrolled in the Software Development Project Management (SDPM) course during the 2021/2022 academic year. SDPM is taken in the second semester of the third year of the Bachelor's Degree in Computer Science. Its contents are related to the so-called "ICT Governance", which helps to train the Computer Engineer in the areas of project management, requirements and software process improvement. Knowledge of these contents will facilitate future work as an engineer in roles such as project manager, ICT department management and/or consultancy work, among others. The course consists of 60 classroom hours, of which 45 hours are theoretical and 15 hours are laboratory practice.

An opt-out recruitment method was employed in

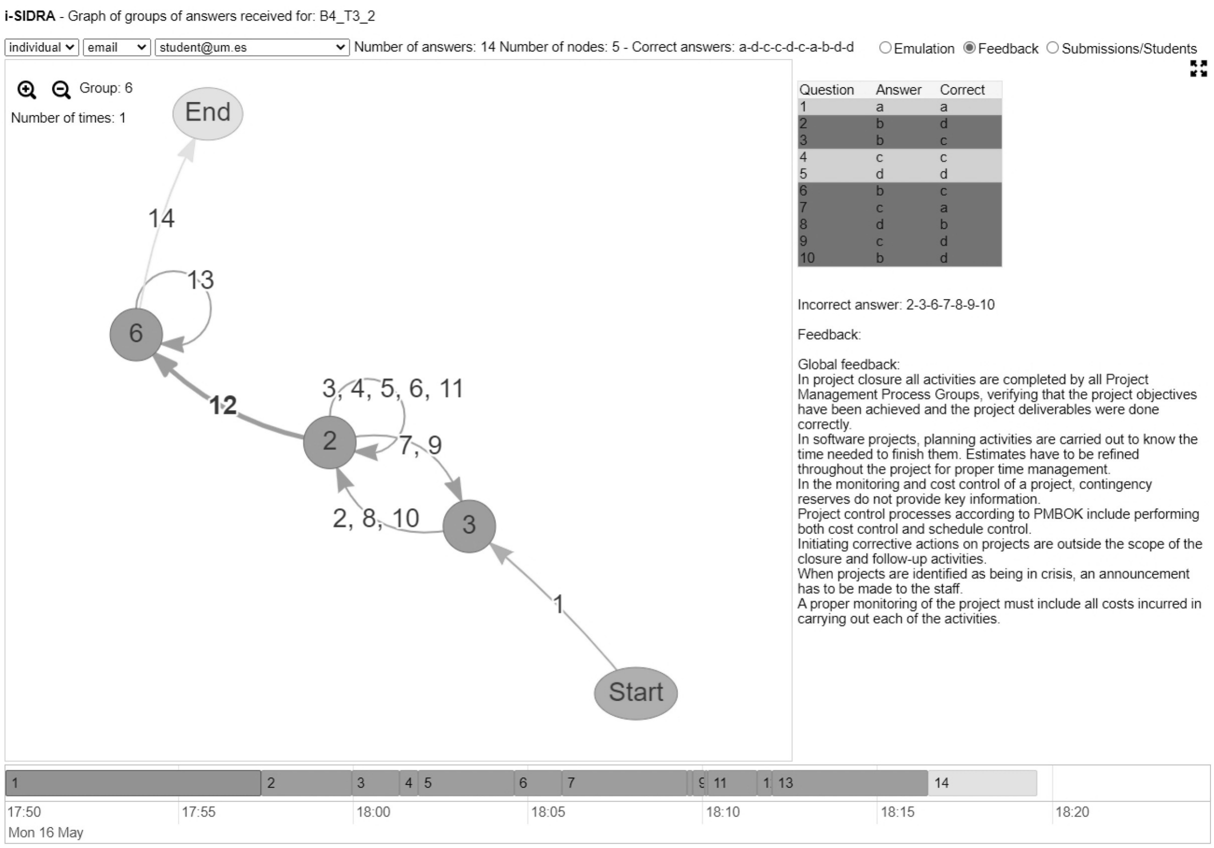

Fig. 2. Evolution by knowledge states and feedback received by the students through the knowledge graph created in a test.

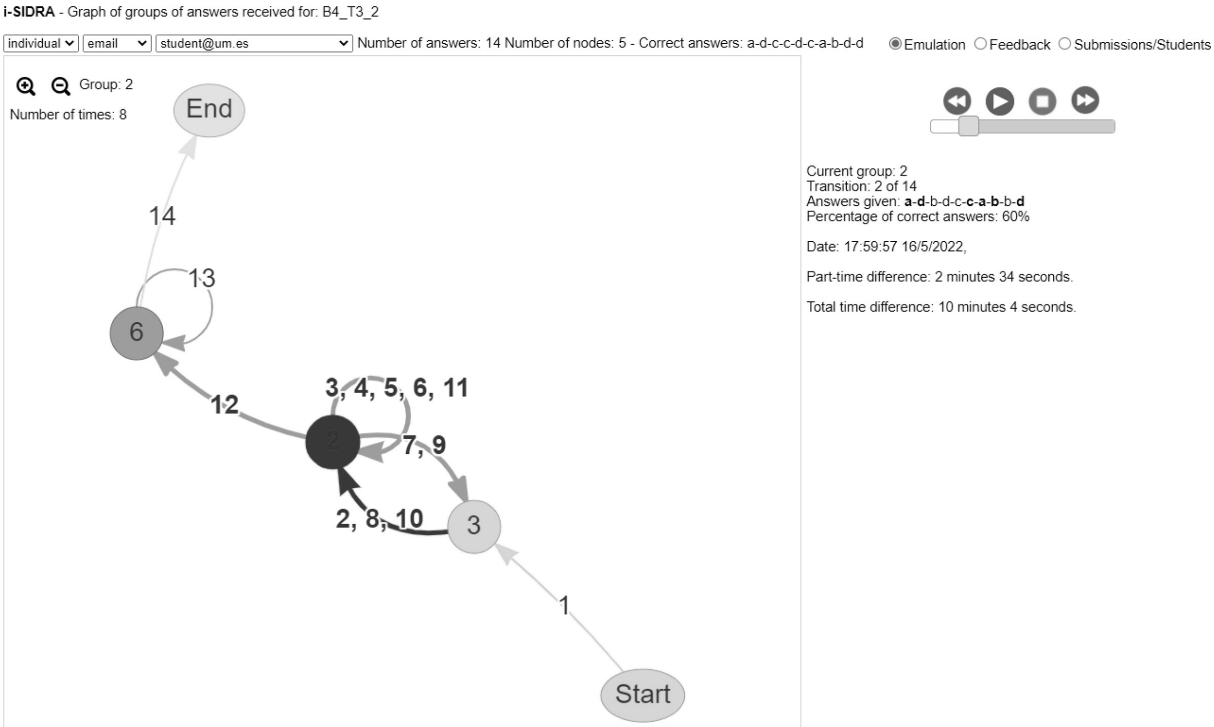

Fig. 3. Playing back the student's transitions through the different states of knowledge.

which participants are given the opportunity to abandon the experiment. Of the 138 participants, 110 (86.3%) were men, and 28 (13.7%) were women. Participants' age ranged from 20 to 24. Most of the participants (86.8%) were in their first enrolment. Ethics committee approval was obtained prior to conducting the study. Sample size calculation was conducted for a two-tailed independent t-test and by using the G*Power 3.1.9 program [50]. With a type I error rate of 0.05, power of 0.80, and an effect size of 0.70, the estimated minimal required sample size for a two-tailed independent t-test was 34 and for a Mann-Whitney-Wilcoxon (two groups) for independent samples was 35.

4.2 Design

For this experiment the students of the course were distributed in 2 groups based on the allocation made by the university's administration and services personnel. Although participants were non-randomized placed in the experiment group and control group, none of the participants had experience in the use of an intelligent feedback system, nor previous knowledge of SDPM. Therefore, we can assume that the study resulted in comparison groups with these balanced characteristics. An experimental group with 80 students who received the SDPM learning using the i-SIDRA tool and a control group with 58 students who received the same contents, but without using i-SIDRA. The tool was used in three sessions of 2 hours in the 12th week of class. At the end of the teaching period, the final exam of the course was given to all the students in order to study the effect of using the system on the students' performance.

In order to study the effect of different types of feedback on the students' performance, two different questionnaire tests were carried out with i-SIDRA in the experimental group, called Test 1 and Test 2. The time for each test was 30 minutes. In each test, two groups were created, one group receiving elaborated feedback and the other group receiving reduced feedback, which was simpler and more concise (see example in Table 2). Six types of elaborated feedback can be used [27]: attribute isolation, topic contingent, response contingent, hints/cues/prompts, bugs/misconceptions and informative tutoring. In this experiment topic contingent is used for the elaborated feedback in which the learner is given information relating contingent to the theme studied, whereas attribute isolation is employed for the reduced feedback, in which information addressing central isolation attributes of the target topic being studied are provided. In the experiments reported in this article, the purpose of feedback in i-SIDRA is not to tell learners what to do next with that information, thus avoiding try again feedback, or what their current knowledge is at that moment, thus avoiding correct response or verification feedback (e.g., using rubrics). Notice that the correct and incorrectly answered questions were never indicated, in order to encourage reflec-

Table 2. Feedback example (translated from the original Spanish feedback)

Reduced feedback – Group 1
– Karlskrona Manifesto. – GUIs overloading may be less efficient. – In MAGERIT, impact is the consequence of a risk.
Elaborated feedback – Group 1
– The Karlskrona Manifesto defines a total of 5 dimensions related to sustainability to promote sustainable development. – An overloaded graphical user interface can lead to higher energy consumption as users spend more time trying to find the information they are looking for. – The MAGERIT methodology has been developed by the Spanish Government's High Council for Electronic Administration with the aim of minimising the risks associated with the use of ICT in public administration.

tion and reasoning, and to avoid falling into the "trial and error" cycle that students may be tempted to fall into when assessed with a set of MCQs. The feedback was prepared following five out of seven principles of good feedback practice to promote self-regulation [42]: (1) "Helps clarify what good performance is" by exemplars used as a valid SWEBOK-based standard for students to compare their answers; (2) "Delivers high quality information to students about their learning" by providing corrective advice, limiting the quantity of feedback so that it can be used effectively and selecting and prioritising areas for improvement; (3) "Encourages positive motivational beliefs and self-esteem" by automated tests with feedback, using feedback oriented to provide information on achievement rather than only about success or failure; (4) "Provides opportunities to close the gap between current and desired performance" by resubmission of responses in MCQ with the aim of achieving the perfect knowledge state; (5) "Provides information to teachers that can be used to help shape the teaching" by playing back and analysing the student's transitions through the different states of knowledge.

4.3 Hypothesis

The following hypotheses were investigated in this work:

- H1. The final exam grades of students using i-SIDRA are higher than those of other students. An independent variable (TeachingMethod) and a dependent variable (Performance as measured by final exam grades) were defined.
- H2. The final grades of students who received elaborated feedback are higher than students who received reduced feedback. An independent variable (FeedbackType) and a dependent variable (Performance as measured by final exam grades) were defined.
- H3. The feedback designed by the instructors and delivered by the i-SIDRA system helps to clarify misunderstandings and identify what students still need to learn. An independent variable (TimePoint) and a dependent variable (IncreaseNumberOfCorrectAnswers, measured by the difference between scores at the beginning and end of the i-SIDRA session) were defined.
- H4. Students are satisfied using i-SIDRA.

4.4 Statistical Analysis

The collected data were analysed using the statistical tool SPSS 24.0 and Microsoft Office Excel 2020. A significance level of 0.050 was used to indicate a statistically significant difference. The Kolmogorov-Smirnov test was used to check whether the study groups had a normal distribution. The t-Student test was employed in case the data of the dependent variable followed a normal distribution. If the data for the dependent variable had a normal distribution, non-parametric tests were used. In particular, with the Mann-Whitney U-test the differences between the medians of two independent groups were compared.

5. Results

Table 3 depicts some descriptive data on the i-SIDRA experiment. First of all, notice that the average time spent by students with reduced feedback (26.610 minutes for Test 1 and 18.740 minutes for Test 2) is higher than the time spent by students with elaborated feedback (16.250 minutes for Test 1 and 15.350 minutes for Test 2). It is worth noting that the test was considered to be finished when the perfect state was reached or when the 30 minutes available expired. This means that students with elaborated feedback finished earlier, i.e. those who reached the perfect state did so in less time than those with clue-based feedback. Note that the number of attempts per student in the tests with elaborated feedback is significantly lower (4.750 for Test1 and 6.290 for Test2) than in the tests with clue-based feedback (19.110 for Exp1 and 8.870 for Exp2). As a result, students with elaborated feedback spent more time reflecting, and did not fall into the "trial and error" cycle.

Statistically significant differences were found in the final grades between the experimental group that used i-SIDRA ($M = 7.650$), and the control group ($M = 6.400$), with $U = 3306$, $p < 0.001$, confirming H1 (Table 4). It is worth noting that while test 1 with elaborated and clue-based feedback was completed by 12 and 9 students (Table 3), respectively, of these only 11 and 6 students (Table 5) took the test. On the other hand, test 2 with elaborated and clue-based feedback was completed by 34 and 38 students (Table 3), respectively, of

Table 3. Classification of the students and the attempts per experiment. PR1: test 1 with elaborated feedback. PR2: test 2 with elaborated feedback. PR3: test 1 with reduced feedback. PR4: test 2 with reduced feedback

Description	Pr1	Pr2	Pr3	Pr4
Students				
Total number of students	12	34	9	38
Number of students reaching the "perfect knowledge state"	9	24	1	26
Number of students who do not reach the "perfect knowledge state"	3	10	8	12
Number of students with only one attempt or submission (students did not receive feedback)	5	11	0	7
Number of students with more than one attempt/submission (students received feedback)	7	23	9	31
Number of students reaching the "perfect knowledge state" with a single attempt (without feedback)	5	10	0	7
Number of students reaching the "perfect knowledge state" with more than one attempt (feedback)	4	14	1	19
Submissions (of test answers)				
Total number of attempts	57	214	172	337
Average time spent in each attempt (minutes)	3.420	2.440	1.390	2.110
Number of attempts per student	4.750	6.290	19.110	8.870
Average time spent per student (minutes)	16.250	15.350	26.610	18.740
Maximum number of attempts made by a student	14	28	33	38
Effectiveness of feedback				
Number of students with more than one attempt (improvement after feedback)	4	17	4	23
Number of students with more than one attempt (worsening after feedback)	2	3	3	5
Number of students with more than one attempt (neutral after feedback)	1	3	2	3

whom 30 and 33 students (Table 5) took the test. No statistically significant differences were obtained in the final score between the elaborated feedback group and the clue-based feedback group in the two experiments, with $T(5,190) = -1.928$, $p = 0.110$ and $U = 443$, $p = 0.474$, rejecting hypothesis H2. However, the mean scores of the elaborated feedback group, $M = 7.590$ in experiment 1 and $M = 8.810$ in experiment 2, were higher than the mean scores of the clue-based feedback group, $M = 7.440$ in experiment 1 and $M = 6.810$ in experiment 2.

No statistically significant differences were reached in the increase of correct questions from the first to the last attempt (variable IncreaseNumberOfCorrectAnswers) between the elaborated feedback group and the reduced feedback group in the two experiments conducted $T(19) = 0.217$, $p = 0.831$ and $U = 699$, $p = 0.542$, respectively, rejecting hypothesis H3. The means of the increments of correct questions for the elaborated feedback group were $M = 0.750$ in experiment 1 and $M = 1.290$ in experiment 2, compared to the means of the reduced feedback group, $M = 0.556$ in experiment 1 and $M = 1.605$ in experiment 2, respectively (Table 6).

Finally, information was obtained about the students' experience with the i-SIDRA tool. For this purpose, after the test, students completed a survey about their participation in this experiment. The online tool Arsyc [19] was used. A questionnaire with a 5-point Likert-type scale (5 = very high; 4 = high; 3 = medium; 2 = low; 1 = very low) was employed. Seven questions were asked: six evaluation questions and one final essay question.

Table 4. Descriptive statistics for performance. "N": number of students; "M": mean; "MD": median; "SD": standard deviation

Group	N	M	MD	SD
Experimental (i-SIDRA)	80	7.650	7.850	9.030
Control	58	6.400	7.050	3.540

Table 5. Descriptive statistics and hypothesis test of the final grades of the group with elaborated feedback and the group with reduced feedback in the two experiments. "N": sample size; "M": mean; "MD": median; "SD": standard deviation; "P": p-value

Group	N	M	MD	SD	P
ExtFeedTest1	11	8.818	9.000	0.468	0.110
RedFeedTest1	6	6.816	7.500	2.519	
ExtFeedTest2	33	7.593	7.800	1.822	0.474
RedFeedTest2	30	7.446	7.700	1.603	

Table 6. Descriptive statistics and hypothesis test of the increase in correct questions from the first to the last attempt of the group with elaborated feedback and the group with reduced feedback in the two experiments performed. "N": sample size; "M": mean; "MD": median; "SD": standard deviation; "P": p-value

Group	N	M	MD	SD	P
ExtFeedTest1	12	0.750	0.000	2.179	0.831
RedFeedTest1	9	0.556	0.000	1.810	
ExtFeedTest2	34	1.294	0.500	2.038	0.542
RedFeedTest2	38	1.605	1.000	2.199	

Table 7. Means, medians and standard deviations of students' perceptions. "M": mean; "SD": standard deviation; "MD": median

Id	Question	M	SD	MD
Q1	In general, are you satisfied with the use of i-SIDRA in the classroom?	3.32	1.20	3
Q2	Do you think that immediate explanations (feedback) helped you in learning?	3.53	1.26	4
Q3	Do you think the immediate explanations (feedback) helped you find the final solution?	3.11	1.41	3
Q4	Do you find that the i-SIDRA system adds incentive/motivates you to study, review and extend subject concepts?	3.37	1.12	3
Q5	Do you like to use i-SIDRA in more subjects?	3.05	1.08	3
Q6	Please indicate your overall assessment of the use of the i-SIDRA platform.	3.32	1.16	3

This survey was completed by 19 students in the experimental group (Table 7).

6. Discussion

Our findings confirm that learning materials with elaborated feedback are processed in greater depth and allowed students to reach the learning objectives earlier. The constructive and supportive information contained in the detailed explanation may better meet the cognitive demand of learners, helping them to integrate prior knowledge with the new information provided by elaborated feedback [27]. Previous studies [51] have shown that elaborated feedback produces less extraneous or extrinsic cognitive load since it decreases the distraction of learners due to elements that do not contribute to the learning experience, but require mental processing.

6.1 Learning Outcomes

Our results confirming H1 were in line with two empirical studies carried out with i-SIDRA in a General and Descriptive Anatomy of the Locomotor System course of the Bachelor of Medicine [21] and the Bachelor of Pharmacy [20]. The effect of intelligent feedback on academic performance, through the ESDNN tool, had already been studied in a programming course and a course on software quality at Bachelor's and Master's level, respectively [22]. In this study, the final exam was divided into two parts: a theoretical part on programming concepts and a practical part on problem solving. Significant improvements were only found in the final exam score of students who used ESDNN compared to those who did not use it. In particular, the enhancement was achieved in the theoretical part, which tested a similar type of knowledge to the type of content of the SDPM course. However, no significant differences were found in the marks for the programming problem solving part. The explanation for this result can be found in the fact that MCQ exams do not allow for a profound learning, particularly in the case of higher levels of Bloom's Taxonomy such as analysis and synthesis, which are essential for programming problems.

The M-OFS system, an evolution of ESDNN, was also evaluated in an experiment in an English grammar course in two Chinese universities [52]. The students who used the tool (experimental group) obtained an average grade of 7.952 (out of 10) compared to 7.130 for the group that did not use it (control group), in line with our results. Similar results were obtained in another experiment applied in two data analysis and programming courses [53].

6.2 Effectiveness of Feedback

Our results show weak indications that elaborated feedback offers better retention and thus a higher Germanic cognitive load, which has been contrasted in previous studies [51] (Table 5). In these previous experiments it has been shown that elaborated feedback produced high motivation and higher levels of Germanic load in learners, namely, it produces more elements that help the learner to consolidate the knowledge acquired during the learning experience in long-term memory [51]. This result can be explained by the fact that detailed information engages learners' deeper cognitive processes, leading to learning gains in subsequent tasks. Further evidence on the effects of elaborated feedback can be found in a meta-analysis that analysed computer-based and non-computer-based learning environments [31].

In this tertiary study, elaborated feedback was found to produce larger effect sizes than simple feedback. Other work that studied feedback exclusively in computer-based learning environments [54–56] found that elaborated feedback is more effective than simple feedback for learning outcomes in university students.

However, elaborated feedback is not necessarily more effective than clue-based feedback [57, 58]. Excessive and redundant feedback could be detrimental to learners' perception and processing of useful information, undermining the positive effect of elaborated feedback. Moreover, feedback may have different effects on learners in different disciplines [59]. Another factor that influences the effectiveness of feedback is the age and prior knowledge level of students [29, 57, 60, 61].

6.3 Increased Knowledge

Our findings with respect to rejecting H3 are inconclusive. One possible explanation for this result is that in a test, pre-set options offer learners the opportunity to easily present answers, which may lead to superficial processing of feedback information, regardless of whether it is elaborated or simple feedback [51]. The effectiveness of feedback depends largely on how it is perceived and interpreted by learners [31, 62, 63].

6.4 Satisfaction

In general, the use of i-SIDRA was rated favourably, highlighting the help offered by the feedback to the student to learn subject knowledge (median of 4), in line with a previous study that used a similar tool (M-OFS) and in which the majority of students (94%) considered that the system helps them to improve their knowledge [52]. These favourable statements for an intelligent feedback system were also obtained in a study with i-SIDRA in a Bachelor of Medicine [21], with an average of 3.900 out of 5, and in a Bachelor of Pharmacy [20], with an average of 4.400 out of 5. In addition, in essay question 6, students highlighted the usefulness of the tool. Some of the comments made by students were: "*It helps to internalise concepts*", "*It reinforces concepts and improves their understanding in a more entertaining way*", "*Use it more often*". Notice that 90.4% of the students who used M-OFS also expressed their willingness to use the system again [52]. These findings are in line with those obtained in a study with i-SIDRA in a Bachelor of Medicine, with an average of 3.900 out of 5, and a Bachelor of Pharmacy, with an average of 4.410 out of 5.

In some studies, both in traditional teaching [64, 65] and in e-learning environments [66], learners perceive elaborated feedback to be more useful than simple feedback. However, these studies could only find positive correlations between perceived usefulness and performance in computer-based teaching [64, 65]. In our study, the questionnaire was anonymous, and no such relationships could be established.

6.5 Limitations

Some limitations of this study should be kept in mind. First, the experiment was conducted in a project management course and learning performance was assessed only by final exam grades. Therefore, caution should be exercised in generalizing our findings to other courses or using other types of instruments to measure learning performance. In addition, the study selected university students as participants, but it should be kept in mind that the effects of feedback may vary with the age of the learner. Moreover, if engineering professionals had been involved, the representativeness of the participants may have been improved. However, as Carver et al. point out [67], the results achieved through empirical studies carried out with students have an impact on the progress of Software Engineering [68]. Controlled experiments with students provide a view of problems that can subsequently be addressed in industrial case studies [69]. Finally, our study presented feedback in textual form. One research direction could be to examine the effects of feedback provided with pictures.

7. Conclusion

The i-SIDRA tool can be used in formative evaluation processes. It allows the student to self-evaluate while constantly reflecting through the personalized feedback received, until the final solution is reached. In turn, the teacher can easily analyse the results of the test to detect possible deficiencies within the group, either globally or individually.

In this experiment, it was observed that the students who used the tool obtained better final grades in the subject. Furthermore, there is no evidence that elaborated feedback offers better long-term retention than reduced feedback. Teachers should be aware of the importance of preparing comprehensive and appropriate feedback to optimize student learning. In general, students are satisfied with the use of the tool in class, highlighting the help it offers them in learning.

One factor that can influence learning is the type of question. MCQs do not favour the acquisition of knowledge and creative thinking of students, because this type of questions allows students to generate answers easily and with less effort than with questions that require an elaborated answer, thus resulting in content superficial processing which negatively affects learning.

There are a number of avenues for future work: (1) extend the study to elaborated response questions; (2) enrich the tool with new methods for the selection of the number of groups; (3) include gamification elements; and (4) automate the process of enriching the answers to create a new knowledge group structure, which currently requires teacher intervention.

Acknowledgments – This research is part of the OASSIS-UMU project (PID2021-122554OB-C32) funded by MCIN/AEI/10.13039/501100011033/ and by ERDF A way to make Europe. This research is also part of the BIZDEVOPSGLOBAL-UMU (RTI2018-098309-BC33) project (from the Spanish Ministry of Science, Innovation and Universities), supported by ERDF.

References

1. M. Mahyoob, Challenges of e-Learning during the COVID-19 Pandemic Experienced by EFL Learners, *Arab World English J.*, **11**(4), pp. 351–362, 2022.
2. M. A. Islam, S. Nur and M. S. Talukder, E-learning in the time of COVID-19: Lived experiences of three university teachers from two countries, *E-Learning Digit. Media*, **18**(6), pp. 557–580, 2021.
3. K. Mukhtar, K. Javed, M. Arooj and A. Sethi, Advantages, Limitations and Recommendations for online learning during COVID-19 pandemic era: Online learning during COVID-19 pandemic era, *Pakistan J. Med. Sci.*, **36**(COVID19-S4), 2020.
4. R. Wood and S. Shirazi, A systematic review of audience response systems for teaching and learning in higher education: The student experience, *Comput. Educ.*, **153**, p. 103896, 2020.
5. W. Beekes, The "Millionaire" method for encouraging participation, *Act. Learn. High. Educ.*, **7**, 2006.
6. F. N. Hussain and K. J. Wilby, A systematic review of audience response systems in pharmacy education, *Curr. Pharm. Teach. Learn.*, **11**(11), pp. 1196–1204, 2019.
7. C. Crockett, C. Finelli, M. DeMonbrun, K. Nguyen, S. Tharayil, P. Shekhar and R. Rosenberg, Common Characteristics of High-quality Papers Studying Student Response to Active Learning, *Int. J. Eng. Educ.*, **37**, pp. 420–432, 2021.
8. Y.-T. Chien, Y.-H. Chang and C.-Y. Chang, Do we click in the right way? A meta-analytic review of clicker-integrated instruction, *Educ. Res. Rev.*, **17**, pp. 1–18, 2016.
9. J. J. López-Jiménez, J. L. Fernández-Alemán, J. A. García-Berná, L. L. González, O. G. Sequeros, J. N. Ros, J. M. Carrillo de Gea, A. Idri and A. Toval, Effects of gamification on the benefits of student response systems in learning of human anatomy: Three experimental studies, *Int. J. Environ. Res. Public Health*, **18**(24), 2021.
10. D. J. Skiba, Got large lecture hall classes? Use clickers, *Nurs. Educ. Perspect.*, **27**(5), pp. 278–280, 2006.
11. P. W. E. Beebe and K. R, Bailey N, Efficacy of a Handheld Response Device to Improve Classroom Interaction and Learning Outcomes. Poster presentation: Emerging Technologies in Nurse Education Conference, *Proceedings of the Emerging Technologies in Nurse Education Conference*, Seattle, USA, pp. 14–15, 2008.
12. N. Efstathiou and C. Bailey, Promoting active learning using Audience Response System in large bioscience classes, *Nurse Educ. Today*, **32**(1), pp. 91–95, 2012.
13. A. Nájera López, E. Arribas Garde, A. Beléndez, J. M. Villalba Montoya, J. Francés and M. J. García Meseguer, Peer evaluation and teaching medical physics using remote response devices (clickers), *Lat. Am. J. Phys. Educ.*, **6**(1), pp. 290–295, 2012.
14. J. E. Caldwell, Clickers in the Large Classroom: Current Research and Best-Practice Tips, *CBE – Life Sci. Educ.*, **6**(1), pp. 9–20, 2007.
15. R. H. Kay and A. LeSage, Examining the benefits and challenges of using audience response systems: A review of the literature, *Comput. Educ.*, **53**(3), pp. 819–827, 2009.
16. Kahoot, https://kahoot.com, Accessed 17 August 2022.
17. Wooclap, https://www.wooclap.com, Accessed 17 August 2022.
18. Socrative, https://www.socrative.com, Accessed 17 August 2022.
19. Arsyc, https://arsyc.com, Accessed 17 August 2022.
20. J. L. Fernández-Alemán, L. López-González, O. González-Sequeros, C. Jayne, J. J. López-Jiménez, J. M. Carrillo-de-Gea and A. Toval, An Empirical Study of Neural Network-Based Audience Response Technology in a Human Anatomy Course for Pharmacy Students, *J. Med. Syst.*, **40**(4), 2016.
21. J. L. Fernández-Alemán, L. López-González, O. González-Sequeros, C. Jayne, J. J. López-Jiménez and A. Toval, The evaluation of i-SIDRA – a tool for intelligent feedback – in a course on the anatomy of the locomotor system, *Int. J. Med. Inform.*, **94**, 2016.
22. J. L. Fernandez-Alemán, D. Palmer-Brown and C. Jayne, Effects of Response-Driven Feedback in Computer Science Learning, *IEEE Trans. Educ.*, **54**(3), pp. 501–508, 2011.
23. J. L. Fernández-Alemán, D. Palmer-Brown and C. Draganova, Evaluating Student Response Driven Feedback in a Programming Course, *Proceedings of the 10th IEEE International Conference on Advanced Learning Technologies*, Sousse, pp. 279–283, 2010.
24. D. Palmer-Brown, C. Draganova and S. Wee-Lee, Guided Learning via Diagnostic Feedback to Question Responses, *SIGCSE Bull.*, **41**(3), p. 362, 2009.
25. J. L. Fernández-Alemán, C. Jayne, A. B. Sánchez, J. M. Carrillo-de-Gea and A. Toval, Neural network-based data analysis for medical-surgical nursing learning, *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego, CA, USA, pp. 6036–6039, 2012.
26. S. Narciss, Designing and Evaluating Tutoring Feedback Strategies for digital learning environments on the basis of the Interactive Tutoring Feedback Model, *Digit. Educ. Rev.*, **23**, pp. 7–26, 2013.
27. V. J. Shute, Focus on Formative Feedback, *Rev. Educ. Res.*, **78**(1), pp. 153–189, 2008.
28. J. Hattie and H. Timperley, The Power of Feedback, *Rev. Educ. Res.*, **77**(1), pp. 81–112, 2007.
29. E. Fyfe, B. Rittle-Johnson, and M. DeCaro, The Effects of Feedback During Exploratory Mathematics Problem Solving: Prior Knowledge Matters, *J. Educ. Psychol.*, **104**, p. 1094, 2012.
30. J. Roelle, K. Berthold and S. Fries, Effects of Feedback on Learning Strategies in Learning Journals: Learner-Expertise Matters, *Int. J. Cyber Behav. Psychol. Learn.*, **1**, pp. 16–30, 2011.

31. R. L. Bangert-Drowns, C.-L. C. Kulik, J. A. Kulik, and M. Morgan, The Instructional Effect of Feedback in Test-like Events, *Rev. Educ. Res.*, **61**(2), pp. 213–238, 1991.
32. Y. Attali, C. Laitusis and E. Stone, Differences in Reaction to Immediate Feedback and Opportunity to Revise Answers for Multiple-Choice and Open-Ended Questions, *Educ. Psychol. Meas.*, **76**(5), pp. 787–802, 2016.
33. Y. Attali and F. van der Kleij, Effects of feedback elaboration and feedback timing during computer-based practice in mathematics problem solving, *Comput. Educ.*, **110**, pp. 154–169, 2017.
34. M. E. Martinez, Cognition and the question of test item format, *Educ. Psychol.*, **34**(4), pp. 207–218, 1999.
35. A. P. Cavalcanti, A. Barbosa, R. Carvalho, F. Freitas, Y.-S. Tsai, D. Gašević and R. F. Mello, Automatic feedback in online learning environments: A systematic literature review, *Comput. Educ. Artif. Intell.*, **2**, p. 100027, 2021.
36. R. Pàmies-Vilà, A. Fabregat-Sanjuan, J. Puig-Ortiz, L. J. Nebot, and A. H. Fernández, Impact of a Gamification Learning System on the Academic Performance of Mechanical Engineering Students, *Int. J. Eng. Educ.*, **5**(A), pp. 1434–1442, 2022.
37. G. G. Krauss, A. Vasquez, J. Sinopoli and L. P. Blake, Upping the average Manipulating peer feedback quantity and its effects on feedback quality, *Int. J. of. Eng. Educ.*, **34**(2B), pp. 674–685, 2018.
38. A. González-Marcos, F. Alba-Elías, and J. Ordieres-Meré, An Online Assessment and Feedback Approach in Project Management Learning, *Proceedings of the CSEDU (2)*, pp. 69–77, 2017.
39. X. Ma, English Teaching in Artificial Intelligence-based Higher Vocational Education Using Machine Learning Techniques for Students' Feedback Analysis and Course Selection Recommendation, *JUCS – J. Univers. Comput. Sci.*, **28**, pp. 898–915, 2022.
40. M. Jaeger and D. Adair, Impact of Computer-Based Feedback Style on Learning Effectiveness in Simulation-Supported Courses, *Int. J. Eng. Educ.*, **36**, p. 10, 2020.
41. D. Loksa, L. Margulieux, B. A. Becker, M. Craig, P. Denny, R. Pettit and J. Prather, Metacognition and Self-Regulation in Programming Education: Theories and Exemplars of Use, *ACM Trans. Comput. Educ.*, **22**(4), 2022.
42. D. J. Nicol and D. Macfarlane-Dick, Formative assessment and self-regulated learning: a model and seven principles of good feedback practice, *Stud. High. Educ.*, **31**(2), pp. 199–218, 2006.
43. C. Ebert, A. Vizcaino, R. Grande, and C. Ebert, Unlock the Business Value of Gamification, *IEEE Softw.*, **39**(6), pp. 15–22, 2022.
44. A. Vizcaíno, F. García, I. G. R. De Guzmán and M. Á. Moraga, Evaluating GSD-Aware: A Serious Game for Discovering Global Software Development Challenges, *ACM Trans. Comput. Educ.*, **19**(2), 2019.
45. J. M. Russell, C. Baik, A. T. Ryan and E. Molloy, Fostering self-regulated learning in higher education: Making self-regulation visible, *Act. Learn. High. Educ.*, **23**(2), pp. 97–113, 2022.
46. J. J. López-Jiménez and J. L. Fernández-Alemán, SIDRA – iOS – App Store, https://apps.apple.com/us/app/sidra/id574060132?l=es&ls=1, Accessed 17 August 2022.
47. J. J. López-Jiménez and J. L. Fernández-Alemán, SIDRA – Android – Google Play, https://play.google.com/store/apps/details?id=cliente.sidra, Accessed 17 August 2022.
48. D. Brown, C. Draganova and S. W. Lee, Snap-drift neural network for selecting student feedback, *Proceedings of the International Joint Conference on Neural Networks*, Atlanta, GA, USA, pp. 391–398, 2009.
49. P. Rousseeuw, Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. Comput. Appl. Math. *J. Comput. Appl. Math.*, **20**, pp. 53–65, 1987.
50. F. Faul, E. Erdfelder, A.-G. Lang and A. Buchner, G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences, *Behav. Res. Methods*, **39**(2), pp. 175–191, 2007.
51. Z. Wang, S.-Y. Gong, S. Xu and X.-E. Hu, Elaborated feedback and learning: Examining cognitive and motivational influences, *Comput. Educ.*, **136**, pp. 130–140, 2019.
52. R. Guo, D. Palmer-Brown, S. W. Lee and F. F. Cai, Intelligent diagnostic feedback for online multiple-choice questions, *Artif. Intell. Rev.*, **42**(3), pp. 369–383, 2014.
53. S. Habte, D. Palmer-Brown, M. Kang and F. Cai, Identifying student group profiles for diagnostic feedback using snap-drift modal learning neural network, *Artif. Intell. Res.*, **5**, 2015.
54. G. Cheng, The impact of online automated feedback on students' reflective journal writing in an EFL course, *Internet High. Educ.*, **34**, pp. 18–27, 2017.
55. B. Finn, R. Thomas and K. A. Rawson, Learning more from feedback: Elaborating feedback with examples enhances concept learning, *Learn. Instr.*, **54**, pp. 104–113, 2018.
56. A. Lachner, C. Burkhart and M. Nückles, Formative computer-based feedback in the university classroom: Specific concept maps scaffold students' writing, *Comput. Human Behav.*, **72**, pp. 459–469, 2017.
57. E. R. Fyfe, Providing feedback on computer-based algebra homework in middle-school classrooms, *Comput. Human Behav.*, **63**, pp. 568–574, 2016.
58. F. van der Kleij, T. J. H. M. Eggen, C. F. Timmers and B. P. Veldkamp, Effects of feedback in a computer-based assessment for learning, *Comput. Educ.*, **58**(1), pp. 263–272, 2012.
59. F. van der Kleij, R. Feskens and T. Eggen, Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes: A Meta-Analysis, *Rev. Educ. Res.*, **85**, 2015.
60. E. R. Fyfe and B. Rittle-Johnson, Feedback Both Helps and Hinders Learning: The Causal Role of Prior Knowledge, *J. Educ. Psychol.*, **108**, pp. 82–97, 2016.
61. U.-M. Krause, R. Stark and H. Mandl, The effects of cooperative learning and feedback on e-learning in statistics, *Learn. Instr.*, **19**(2), pp. 158–170, 2009.
62. D. Ilgen, C. Fisher and M. Susan, Consequences of individual feedback on behavior in organizations, *J. Appl. Psychol.*, **64**, pp. 349–371, 1979.
63. C. F. Timmers, A. Walraven and B. P. Veldkamp, The effect of regulation feedback in a computer-based formative assessment on information problem solving, *Comput. Educ.*, **87**, pp. 1–9, 2015.
64. B. Harks, K. Rakoczy, J. Hattie, M. Besser and E. Klieme, The effects of feedback on achievement, interest and self-evaluation: the role of feedback's perceived usefulness, *Educ. Psychol.*, **34**(3), pp. 269–290, 2014.
65. K. Rakoczy, B. Harks, E. Klieme, W. Blum and J. Hochweber, Written feedback in mathematics: Mediated by students' perception, moderated by goal orientation, *Learn. Instr.*, **27**, pp. 63–73, 2013.

66. J.-W. Strijbos, S. Narciss and K. Dünnebier, Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency?, *Learn. Instr.*, **20**, pp. 291–303, 2010.
67. J. Carver, L. Jaccheri, S. Morasca and F. Shull, Issues in using students in empirical studies in software engineering education, *Proceedings of the 5th International Workshop on Enterprise Networking and Computing in Healthcare Industry (IEEE Cat. No.03EX717)*, pp. 239–249, 2003.
68. I. Salman, A. T. Misirli and N. Juristo, Are Students Representatives of Professionals in Software Engineering Experiments?, *Proceedings of the 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, pp. 666–676, 2015.
69. E. Arisholm, L. C. Briand, S. E. Hove and Y. Labiche, The impact of UML documentation on software maintenance: an experimental evaluation, *IEEE Trans. Softw. Eng.*, **32**(6), pp. 365–381, 2006.

Juan José López-Jiménez, is currently a PhD student and collaborator in the Software Engineering Research Group, University of Murcia, Spain. He received the technical engineering degree in system informatics and a master's degree in computer science from the University of Murcia, Spain, in 2012 and 2015, respectively, and the degree in computer science from the University of Castilla-La Mancha, Spain, in 2014. In addition, he obtained a master's degree in Industrial Automation from the Catholic University of Avila, Spain, in 2014.

José Alberto García-Berná, is currently a research assistant at the Software Engineering Research Group, Faculty of Computer Science, University of Murcia. He received the BSc in telecommunication engineering from the Technical University of Valencia, and the PhD degree in computer science from the University of Murcia, Spain, in 2021. His research interests are requirements engineering and project management, in particular green software engineering and energy efficiency in information systems.

Ambrosio Toval, is currently a Full Professor with the University of Murcia. He received the BSc degree in mathematics from the University Complutense of Madrid, Madrid, Spain, in 1983, and the PhD degree (cum laude) in computer science from the Technical University of Valencia, Valencia, Spain, in 1994.

José Luis Fernández-Alemán, is currently an Associate Professor with the Department of Informatics and Systems, University of Murcia, Spain. He received the BSc degree (Hons.) in computer science and the PhD degree in computer science from the University of Murcia, Spain, in 1994 and 2002, respectively.

Ali Idri, is currently a Professor with the Computer Science and Systems Analysis School (ENSIAS), University Mohammed V of Rabat. He received the DEA (Master) and Doctorate of 3rd Cycle degrees in computer science from the University Mohammed V of Rabat, Rabat, Morocco, in 1994 and 1997, respectively, and the PhD degree in cognitive computer sciences from the University of Quebec, Montreal, QC, Canada, in 2003.