Received: 16 January 2017 | Accepted: 23 July 2017
DOI: 10.1002/cae.21869

RESEARCH ARTICLE

WILEY

Video podcast and illustrated text feedback in a web-based formative assessment environment

Emilio Letón[1] | **Elisa M. Molanes-López**[2] | **Manuel Luque**[1] | **Ricardo Conejo**[3]

[1] Department of Artificial Intelligence, Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain

[2] Department of Statistics and Operations Research, Universidad Complutense de Madrid (UCM), Madrid, Spain

[3] Department of Languages and Computer Science, Universidad de Málaga (UMA), Málaga, Spain

Correspondence
Emilio Letón, Department of Artificial Intelligence, Universidad Nacional de Educación a Distancia (UNED), Madrid, 28040 Spain.
Email: emilio.leton@dia.uned.es

Abstract

It is well known that computer-based formative assessment and timely feedback enhance effective student learning but there is still a debate about what type of feedback should be given, being the text-based the most used feedback in practice. Although the use of video content as a learning resource has recently increased in both educational and non-educational contexts, there is very little research on its effectiveness as assessment feedback and the existing results are contradictory. For that reason, we have combined in this work the use of a web-based formative assessment system (Siette) where we have integrated a specific type of video podcast ("modular teaching mini-videos": MTMs) and equivalent illustrated text. We have carried out two experiments with a twofold purpose, first to compare the effect of video feedback versus correct response feedback alone and secondly to compare the effect of video feedback versus equivalent illustrated text feedback. In the context of our research, a Statistic Course at university level, our results show, as expected, that there is a statistically significant positive effect in favor of video feedback over correct response feedback alone. Surprisingly, in contrast to our expectations, based on our context of acquisition of procedural knowledge, there is a statistically non-significant effect of video feedback versus equivalent illustrated text feedback. From the students' satisfaction survey, there is not statistically significant differences in the overall score of the activity based on video podcast feedback or equivalent illustrated text, which is an indication that we have really obtained equivalent feedback materials.

KEYWORDS

formative assessment, illustrated text feedback, learning effectiveness, media in education, video feedback

1 | INTRODUCTION

In the literature, there is a consensus that assessment is essential to student learning in higher education (see, for example, ref. [39]) and that feedback is a central aspect of the assessment process (see, among other ref. [2,5,35,22]). Additionally, it is known that high quality and timely feedback for formative, sumative, and continuous assessment enhance effective student learning (see, for example, ref. [24,28,13]). Despite many authors consider that new information technologies [7,12,21,29,41,43,58] are very helpful to achieve timely feedback, there is a lack of consensus about the type of feedback that should be given [26]. Although in recent years, there has been an increasing interest in the use of video content as a learning resource [30], there has been very little research investigating video-based assessment feedback, being the text-based assessment feedback the most used as ref. [26] and [39] point out. The vast majority of the researches that study video-based assessment feedback, make a comparison with text-based assessment feedback that is not equivalent in the volume of detail of information. For that reason, the purpose of our work is twofold. First, to study the effect of video feedback in terms of student academic performance and, second, to compare it with the effect of feedback based on *equivalent illustrated text* considering that it refers to a written version of the video where the colloquial oral explanations are replaced by equivalent written forms and the handwritten drawings and expressions on the video to strengthen the oral explanation are replaced by equivalent digital ones. Our context of research will be a formative assessment environment on a Statistic Course at university level.

The structure of the paper is as follows. Section 2 describes the feedback effect in formative assessment environments, analyzing current findings on the use of video and text as learning resources. We describe the methodology used in our two experiments in Section 3, and analyze their results in Section 4. Finally, we give our conclusions and describe our future work in Section 5.

2 | FEEDBACK EFFECT IN A FORMATIVE ASSESSMENT ENVIRONMENT

Feedback is defined as any piece of information that is given to the student after one of the student's actions. Feedback is thus a "consequence" of performance. Authors in ref. [24] provided a conceptual analysis of feedback and reviewed the evidence related to its impact on learning and achievement. According to these authors, teachers should consider three dimensions to ensure that the feedback given to students is effective: (i) providing information about the attainment of learning goals related to the task or performance; (ii) providing information about progress and/or about how to proceed; and (iii) providing information that leads to greater opportunities for learning, including enhanced challenges.

Formative assessment provides feedback to students at the same time as they are engaged in a learning activity. Authors in ref. [3,4] established a relationship between formative assessment and well-founded pedagogical theories that have their roots in a constructive perspective, where learners are seen as actively constructing knowledge and understanding through cognitive processes (Piaget's cognitive development) in a social and cultural context (Vygotzky's zone of proximal development). According to these authors, one of the key features of formative assessment is the feedback it provides to the teacher and students. In this work, we will focus on student feedback.

In a formative assessment environment, feedback can be classified according to timing and content [49]. On one hand, based on timing, feedback may be (i) *immediate*, when questions are posed one by one and feedback appears after each answer is given; or (ii) *delayed*, when feedback is given after the test has been completed. On the other hand, based on content, feedback can be classified as (i) *Knowledge of Response*, usually known as KR, where the feedback consists of merely stating whether the given response was correct or incorrect; (ii) *Knowledge of Correct Response*, usually known as KCR, where the feedback contains not only the correctness of the response but also the correct answer if an incorrect response was given; and (iii) *Elaborate Feedback*, known as KCR + EF, where an extra explanation is provided in addition to KCR feedback.

2.1 | Video podcast as a learning resource

In recent years, there has been an increasing level of interest in the use of video podcast as a learning resource [30]. Two factors have contributed to this interest. First, there is an increasing demand for Massive Open Online Courses (MOOCs), such as Coursera or Udacity, or other web educational resources such as Khan Academy, which are mostly based on audio-visual material [56]. Second, the bandwidth currently available for Internet connections has increased, and certain initial problems with using video files have been solved, as noted in earlier works [1], which has increased the amount of multimedia content [59].

It is known that video podcasts, if designed properly, can enhance learning compared to more traditional teaching methods (see, for example, ref. [30,36,38,64], among others). In fact, some factors that correlate with effective learning are:

1. *Motivation*: affective and cognitive student attitudes toward video podcasts are predominantly positive [9,30].
2. *Interactivity*: in the sense that the user can access a specific moment of video content at any time [65], has to answer a set of questions after viewing the video podcast [47], or has to use software programs to supplement the instruction delivered through the streamed media [23]. The effectiveness of interactivity is congruent with the constructivist theory of learning [27].
3. *Short duration*: some authors noted that the use of long videos is commonly associated with passive watching [50,61], and that the use of short videos is associated with interactive tasks [17,25,65].

2.2 | Comparison of video and text format for learning

There are not many studies that compare the use of video and text format for learning purposes and the results are contradictory. Some studies in non-educational contexts have shown that the text is superior, for instance, for memorization of news or political broadcasts [19,20,57], for understanding of an organizational mission statement [21], and for recalling of television commercials [16]. Some other studies in educational contexts have shown the opposite in areas connected with abilities, for instance, for practical skills [15], for golf swing [18], for tying nautical knots [48], and for laparoscopic knot tying [63], or in areas that require the acquisition of procedural knowledge, for instance, in Biology [32] and Statistics [14]. There are also studies that do not find significant differences [40]. These discrepancies can be explained from a theoretical perspective because of several reasons as pointed out by ref. [40]: (i) the dynamic visual presentation of videos is specially beneficial for topics that require a procedural learning rather than a declarative learning; (ii) the inclusion of multimedia effects is beneficial according to dual-coding theory [44] and; consequently, it is expected that both "text combined with pictures" and video are superior to text alone, but, nevertheless, it is not so clear if video is superior to "text combined with pictures" or the other way around; (iii) the modality principle would predict the superiority of "pictures combined with spoken short text" over "pictures combined with written short text" [33] but not for the case of long texts.

As ref. [40] pointed out, a possible reason for the above-mentioned differences is how much control the recipient can apply over the information processing. Textual presentation allows learners to reread relevant passages, skip unimportant ones, and adjust the reading pace to individual cognitive needs. This might be a drawback of long video podcasts or those simultaneously presented to a group of learners where the video is played from the beginning to the end [30]. For that reason, authors in ref. [40] decompose the activities involving text and video processing into micro- and macro-levels. Micro-level activities are those related to the control of information processing, for instance, using navigation through single words, sentences, or short paragraphs, that is, rereadings and lookbacks, on a local text level, or using play, stop, forward and backward buttons on video. Macro-level activities include the use of top-down organizers such as tables of contents or indexes to locate, relate, and compare relevant parts of text or video. Authors in ref. [40] conclude that micro-level activities correlate more with better effective learning using video. Additionally, the above-mentioned comparisons are sometimes unfair because the information content is not equivalent between the two formats. In other cases, the advantage of video instruction is just a matter of user preference, but it is not correlated with better results [14].

2.3 | Purpose of the study

In the literature, there are several studies that consider video podcast as feedback. For example, ref. [8,6,13,26,39,51,60] show that, based on students' satisfaction surveys, the feedback experience can be enhanced by the use of video. However, none of the above studies have conducted an intervention study, similar to ref. [62], to examine whether the student performance is significantly improved when they receive video feedback and whether this improvement is comparable with that achieved when other forms of feedback are provided.

This intervention study with multimedia feedback was suggested to be done in ref. [30,53], but as far as we know it was never done. The study that we have performed is in a formative assessment environment based on immediate elaborate feedback (following ref. [37,52,53]), where we first have compared KCR with KCR + video podcast and then KCR + video podcast with KCR + equivalent illustrated text. These types of feedback cover the three dimensions of effective feedback described by ref. [24]. The type of video podcast used here is short and interactive to allow the use of the micro-level activities considered by ref. [40].

3 | METHODOLOGY

3.1 | Participants

The participants in this study were students enrolled in a Statistics Course, one of the subjects of the Bachelor's Degree in Telecommunications Engineering at University Carlos III of Madrid. The specific statistical topic covered in this research was "the cumulative distribution function (cdf) of random discrete variables."

3.2 | Design

The experimental design was guided by our two initial hypotheses in the domain of Statistics learning at university level:

1. H1A: Using *KCR + video podcast* as elaborate feedback in a web-based formative assessment environment is effective for learning.
2. H1B: Using *KCR + video podcast* is more effective than *KCR + equivalent illustrated text* when used as elaborate feedback in a web-based formative assessment environment.

To analyze these hypotheses, we designed two separate experiments (1A and 1B, respectively), which were conducted in different academic years. For each experiment, students were randomly assigned to one of the following groups: control and experimental. The design of the two experiments can be seen in Table 1.

The experiments were conducted in three phases: pretest, training, and posttest. After the third phase, we collected student feedback through an anonymous survey.

TABLE 1 Experimental design

Experiment	Group	Feedback
1A	Control	KCR
	Experimental	KCR + video podcast
1B	Control	KCR + video podcast
	Experimental	KCR + equivalent illustrated text

3.3 | Learning resources

In this section, we describe the web-based assessment environment Siette, the two resources used as elaborate feedback (a specific type of video podcast known as MTM and equivalent illustrated text) and the integration of each of these types of EF into Siette.

3.3.1 | Siette

Siette [10], is a domain-independent web-based formative assessment system that has been developed at the University of Málaga (Spain). It is available at http://www.siette.org. This system implements the classical test theory and Item Response Theory (IRT). Items (questions) are attached to different topics in the curriculum of a hierarchically structured subject. Siette supports the complete assessment process: creation of an item bank, definition of tests, delivery through a web interface, and data collection and analysis. Additionally, it supports different types of questions, including multiple-choice, single- and multiple-answer and constructed response question based on short answers that are corrected according to regular expression patterns. All questions in Siette are defined in HTML and can therefore be displayed in a web browser. This also allows the inclusion of multimedia resources and illustrated text as elaborate feedback related to the student's responses. For the reader interested in the evolution and current state of Siette, we refer to ref. [11]. In this work, we have included either MTMs or equivalent illustrated texts as EF in Siette to be displayed only when an incorrect answer is given by the student.

Another interesting feature of Siette is the possibility of using item templates that allow us to create generative questions, that is, questions that are instantiated at run time with random values of some parameters. A total of 77 multiple-choice questions (MCQs), most of them generative, were created for the experiments.

Figure 1 illustrates what happens if a student gives an incorrect answer: Siette shows the KCR feedback using a red tick for the correct answer and a red cross for the wrong answer. It is worth mentioning that if the question were answered correctly, Siette would only mark the correct answer with a green tick. In Figure 2 we show two versions of the same generative MCQ.

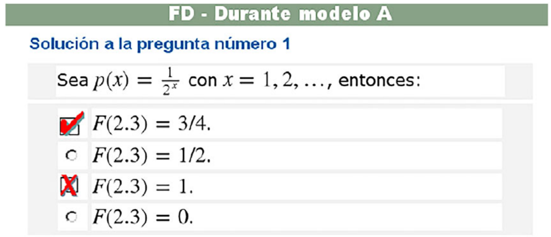

FIGURE 1 A MCQ with KCR in Siette

3.3.2 | Modular teaching mini-videos

The type of video podcasts included in the feedback are "modular teaching mini-videos" (MTMs), which are characterized by their modularity, interactivity, and short duration [34]. They are similar to problem-based video podcasts described by ref. [31], although MTMs are based on *minimalist slides* (MSs). The main features of the set of MSs used in an MTM are that the number of lines on each slide is less than or equal to 7, there is blank space to write on each slide and the slides are available for the student to print them (see ref. [34], for more details). Additionally, the image of the teacher appears in the lower corner of the video, according to the guidelines recently proposed by [54]. The left part of Figure 3 shows a minimalist slide. In designing the MTMs we have taken into account the guidelines described in previous successful works that used video podcast, paying special attention to the following features:

1. *Modularity*. The teacher structures the content of the MTM to synthesize a theoretical concept or to solve a simple exercise. Instead of creating a single video podcast, the content is divided into several separate chunks with an independent URL to be used. Thus, modularity allows the MTM to be reused for different subjects or courses and, in this case, to integrate it into Siette.
2. *Interactivity*. The teacher tries to promote interactivity with students in three ways: (i) by asking them at least one question during the MTM; (ii) by asking him/her to stop the MTM and encouraging them to make a summary; and (iii) by filling the MS to encourage students to do the same on their printed MS. Recently, authors in ref. [36] proved that the learning results are better when the teacher fills the MS than when the slide is already filled.
3. *Short duration*. The duration of an MTM is between 5 and 10 min. This implies that the content of a semester course with six European Credit Transfer System (ECTS) (approximately 6 chapters) could be structured into approximately 120 MTMs (20 MTMs per chapter, so 1 MTM = 1/20 ECTS). The ECTS is a standard used in the European Union to measure the hours of workload by students in higher education (1 ECTS is equivalent to 25–30 hr of study).

Pretest (left panel)	Posttest (right panel)
Let $p(x) = \frac{1}{2^x}$, with $x = 1, 2, \ldots,$ then:	Let $p(x) = \frac{1}{2^x}$, with $x = 1, 2, \ldots,$ then:
1. $F(2.3) = 3/4$.	1. $F(0.5) = 3/4$.
2. $F(2.3) = 1/2$.	2. $F(0.5) = 0$.
3. $F(2.3) = 1$.	3. $F(0.5) = 1/2$.
4. $F(2.3) = 0$.	4. $F(0.5) = 1$.

FIGURE 2 Example of an MCQ used in the pretest (left panel) and posttest (right panel) for the MTM "Example 1 of a cdf of a random discrete variable"

The 4 MTMs that have been used are:

1. Cdf of a random discrete variable: http://minivideos.uc3m.es/wmv/Funcion_de_probabilidad.wmv
2. Example 1 of a cdf of a random discrete variable: http://minivideos.uc3m.es/wmv/Ej_1_vad.wmv
3. Example 2 of a cdf of a random discrete variable: http://minivideos.uc3m.es/wmv/Ej2_vad.wmv
4. Probability of intervals of a random discrete variable: http://minivideos.uc3m.es/wmv/Prob_intervalos_vad.wmv

Figure 3 shows an empty minimalist slide of the MTM "Example 2 of a cdf of a random discrete variable" (left panel) and minute 01:02 of the same MTM with the minimalist slide filled by the teacher (right panel).

3.3.3 | Equivalent illustrated text

As an alternative presentation of the MTM, an equivalent illustrated text was included in the feedback. The process to make the equivalent illustrated text was as follows: first, we took the transcription of the whole MTM and then we replaced the colloquial oral explanations with equivalent written forms and the handwritten drawings or expressions with equivalent digital ones.

By way of illustration, Figure 4 shows the English transcription of the first part of the MTM "Example 2 of a cdf of a random discrete variable" corresponding to minutes 00:00–01:02, and Figure 5 the equivalent illustrated text of this part.

3.3.4 | Integrating the elaborate feedback into Siette

The integration into Siette of the elaborate feedback was straightforward. First, the illustrated texts were integrated as HTML content. Second, the integration of MTMs was also quite simple as embedded code, even though Siette, and MTMs were hosted by different sites. From the students' perspective, everything functioned smoothly as a single site. It is important to notice that when the students give a wrong answer to an MCQ in Siette they receive just one feedback message either via video or illustrated text depending on the experimental group they were assigned to.

3.4 | Assessment tools

The pretest was designed to estimate the students' baseline knowledge before the training. The assessment of this first phase consisted of 10 MCQs, each with four possible options and only one correct answer. For each question, 1 point was given for the correct answer, $-1/3$ for an incorrect answer and

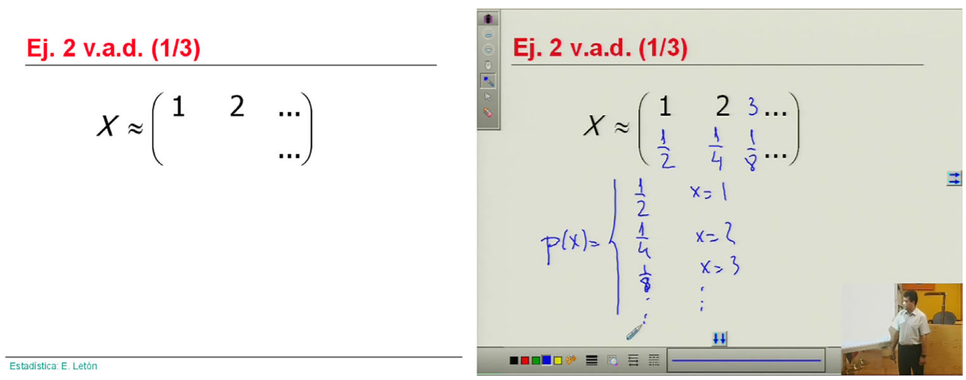

FIGURE 3 Left panel: A minimalist slide of the MTM "Example 2 of a cdf of a random discrete variable". Right panel: Minute 01:02 of the same MTM with the minimalist slide filled

Example 2 of distribution function of a discrete r.v. (part 1 of 3)

In this example we will work with an infinite discrete random variable X. This random variable takes the values $1, 2, 3$, etc. Its support is $1, 2, 3, \ldots$, and takes the value 1 with probability 1/2, the value 2 with probability 1/4, the value 3 with probability 1/8, and so on.

This is the same as saying that $p(x) = 1/2$ if $x = 1$, $p(x) = 1/4$ if $x = 2$, $p(x) = 1/8$ if $x = 3$ and so infinitely since it is an infinite random variable.

This function is a probability function because it effectively satisfies the two conditions: it lies between 0 and 1, and its sum over the support is 1.

FIGURE 4 Transcription of MTM corresponding to minutes 00:00–01:02

0 for a blank response. Since the aim of the training phase is that the students improve their baseline knowledge, there was no assessment for this phase. The posttest was used to assess the students' knowledge after the training. This test contained 10 MCQs, identical to those used in the pretest but with different numerical values to ensure the same level of difficulty for both the pretest and posttest. Figure 2 shows an example of one of these questions.

After the third phase, we collected students' opinion in a survey carried out voluntarily and anonymously in each group. Table 2 shows the items comprising the survey. The first column refers to the questions, and the second column indicates the type of response of each question: a 0–10 scale, yes/no or open. Note that in experiment 1A "activity" means KCR feedback for the Control group and KCR + video podcast feedback for the Experimental group. In the case of experiment 1B, "activity" refers to KCR + video podcast feedback for the Control group and KCR + equivalent illustrated text for the Experimental group. For both experiments "testing tool" refers to Siette in both groups.

3.5 | Procedure

On the Statistics Course, one of the subjects of the Bachelor's Degree in Telecommunications Engineering at University Carlos III of Madrid, the common practice is to have a 90-min *theoretical lecture* (with slides and a blackboard) followed by a 90-min *practical lecture* (where exercises are solved with or without computers) in the same week. Details of the experiments 1A and 1B are given below.

3.5.1 | Experiment 1A

In this experiment, a total of 44 students attended a theoretical lecture given by one instructor (one of the authors of this paper), corresponding to the contents of the 4 MTMs introduced in Section 3.3. In this experiment, the associated practical lecture took place in computer labs. The students were not informed beforehand about the content of the experiment in order to avoid any interferences in their normal learning process. The 44 students were randomly assigned to two different computer labs: one for the control group and one for the experimental group. This assignment was single-blinded for the students. The experiment was conducted in three phases: pretest, training, and posttest. The students were informed about these phases at the beginning of the experiment and received written instructions before each phase. In these three phases, the students had to answer online MCQs in Siette, which were presented one at a time and randomly. The scoring for right answers, wrong answers, and blank responses was given in the instructions. Although the activity was conducted online, the students received enough

Example 2 of distribution function of a discrete r.v. (part 1 of 3)

In this example we consider an infinite discrete random variable (r.v.) X, that takes values in the support $\{1, 2, 3, \ldots\}$ and with probabilities $\left\{p(1) = \frac{1}{2},\ p(2) = \frac{1}{4},\ p(3) = \frac{1}{8}, \ldots\right\}$, that is, with probability function given by

$$p(x) = \begin{cases} \frac{1}{2}, & x = 1 \\ \frac{1}{4}, & x = 2 \\ \frac{1}{8}, & x = 3 \\ \vdots & \vdots \end{cases}$$

In fact, $p(x)$ is a probability function because it satisfies the two required conditions, it lies between 0 and 1 ($0 \leq p(x) \leq 1$, $x = 1, 2, 3, \ldots$) and its sum over the support is equal to 1 ($p(1) + p(2) + p(3) + \ldots = 1$).

FIGURE 5 Equivalent illustrated text corresponding to minutes 00:00–01:02

TABLE 2 Survey administered to students

Question	Type
1. What is your overall rating of the activity?	0–10
2. What is your overall rating of the testing tool?	0–10
3. Would you like to have similar activities in the lab?	Yes/No
4. Would you like to have similar activities available at home?	Yes/No
5. Could you please provide general comments on what you liked best	Open
6. Could you please provide general comments on what you liked least	Open

blank paper to do intermediate calculations before answering the questions. Moreover, to motivate students, they could receive as much as one extra point to their grade depending on their performance in the third phase. During the whole experiment a tutor was present and merely invigilated.

In the pretest phase, all students (control and experimental groups) answered the same 10 MCQs in Siette in 15 min. In this first phase, Siette did not provide the correct answer or indicate whether the answer given was correct.

In the training phase, all students received training based on 16 generative MCQs (that is, generated at run time from 16 item templates) in Siette corresponding to the topics of the 4 MTMs. In the control group, only the correct answer was given, that is, they received KCR feedback. In the experimental group, it was also suggested that they should watch an MTM if their answer was incorrect, that is, they received KCR + video podcast feedback. In this training phase, both groups of students completed a formative assessment with Siette for 25 min. The formative assessment was constructed as an endless loop of 4 MCQs, one for each MTM of study. If they finished in the 25 min allotted, they took another loop of 4 MCQs. To maintain silence in the experimental group while the students listened (and watched) to the MTMs, they received headphones at the beginning of this phase. The way of proceeding in the training phase implies that on average, the experimental group answered fewer questions than the control group because they spent part of their time watching the MTMs.

In the posttest phase, all students answered the same 10 MCQs in Siette in 15 min. These MCQs were completely analogous to those of the pretest phase, with the same wording and just different numerical values (see Figure 2).

3.5.2 | Experiment 1B

In this second experiment that took place in the following academic year, a total of 166 students from four academic groups received a theoretical lecture with the same contents and the same instructor as in experiment 1A, although at different schedules. A random single-blinded assignment was done in six different computer labs: three for the control group and three for the experimental group. This experiment had the same three phases and MCQs as experiment 1A, with the only difference that in the training phase the control group received EF with MTMs, "KCR + video podcast (MTM)" and the experimental group received EF with equivalent illustrated text "KCR + equivalent illustrated text." Again, the students of the "KCR + video podcast" group received headphones at the beginning of the experiment.

3.6 | Analysis of data

The collected data are analyzed using the *R* statistical software. As an indication of the assessment reliability, we use the standard Cronbach's alpha and to measure the effect sizes, we use the Cohen's d. To compare differences in terms of quantitative variables, we use parametric tests (Student's *t*-test or paired *t*-test) if the assumptions of normality and homocedasticity are met, and nonparametric tests (Mann–Whitney–Wilcoxon test or paired Mann–Whitney–Wilcoxon test) if any of these assumptions are not met. We assess the normality assumption through the Shapiro–Wilk statistic and the homocedasticity condition through the Levene test. In order to consider simultaneously several explanatory variables, we use linear regression fit. To compare differences in terms of qualitative variables, we use the Chi-squared test. For all these tests, a significance level of 5% is considered.

4 | RESULTS

In this section, we show the results obtained from the analysis of the two experiments carried out.

4.1 | Experiment 1A

A total of 44 students attended the theoretical lecture as a single group and 35 of them attended the practical lecture: 19 in the Control group and 16 in the Experimental group. The following basic variables were recorded during the course of the experiment:

1. *PreTestScore*: the score obtained by the student in the pretest with a maximum value of 10 points. Due to the substraction of points for questions answered incorrectly the *preTest Score* variable can be negative.
2. *PostTestScore*: the score obtained by the student in the posttest with a maximum value of 10 points. Like the *pre Test Score* variable, due to the substraction of points for incorrect answers, the *preTest Score* can be negative.
3. *NumberOfQuestions*: the total number of questions answered by the student in the training phase. Taking into account that both groups participated in the training

phase for exactly the same duration (25 min), it is expected that the number of questions answered is much higher in the Control group than in the Experimental group.

4. *NumberOfIncorrectAnswers*: the total number of questions incorrectly answered by the student in the training phase.

From these basic variables recorded, we calculated the following three variables:

1. *PercentOfIncorrectAnswers*: the percentage of incorrect answers during the training phase defined as number of Incorrect Answers/number of Questions × 100.
2. *Diff Score*: the absolute learning gain defined as *postTest Score–preTest Score*.
3. *CorrectedLearningGain*: the relative learning gain defined as diffScore/(10−preTest Score).

Both *diff Score* and *correctedLearningGain* are considered to measure the "learning increase score" achieved by the student after the training phase.

Table 3 presents the average results of these variables for the students in the Control and Experimental groups with the standard deviations in parentheses. From this table, we observe that both groups had a *preTest Score* lower than zero, which can be explained by the fact that students do not usually study on a daily basis. It is worth remembering that the students did not know anything about the pre- and posttests involved in this experiment for evaluating their knowledge. The standard Cronbach's alpha of the *preTest Score* was 0.62, not a very high value, which can be explained by the low results obtained. Regarding the *postTest Score*, the standard Cronbach's alpha was equal to 0.89, indicating a highly reliable test. As expected, the number of questions answered by the Control group was much higher than that answered by the Experimental group. Figure 6 shows a graphical summary of the *preTest Score* and *postTest Score* variables and Figure 7 shows a graphical summary of the *diffScore* and *correctedLearningGain* variables.

First, we can consider that the baseline knowledge of each group was the same because there is no statistical significance in the *preTest Score* variable (Mann–Whitney–Wilcoxon, p-value = 0.920). In other words, we can assume that both groups are homogeneous, indicating that the random assignment was successful in the sense that both groups are comparable before the training session.

Second, we measure the "learning increase score" in terms of the *diffScore* and *correctedLearningGain* variables. Regarding the *diffScore* variable, there is a statistical difference in the Experimental group (paired t-test, p-value = 0.027) but not in the Control group (paired t-test, p-value = 0.111). The effect sizes (measured in terms of Cohen's d) were 0.613 for the Experimental group and 0.384 for the Control group, which are considered medium and small effects, respectively. Regarding the *correctedLearningGain* variable, no statistical differences are observed either in the Experimental group (paired t-test, p-value = 0.061) or in the Control group (paired t-test, p-value = 0.146). The effect sizes (measured in terms of Cohen's d) were 0.507 for the Experimental group and 0.348 for the Control group, which are considered medium and small effects, respectively.

If we now take into account not only the *Group* (0 for the Control group and 1 for the Experimental group) but also the percentage of incorrect answers during the training phase, *percentOfIncorrectAnswers*, we obtain the following linear regression fits.

Considering *diffScore* as the dependent variable, we obtain: diffScore = 3.600 + 2.145 × Group−0.102 × percentOfIncorrectAnswers, with a p-value = 0.018 associated with *Group* and a p-value <0.001 associated with *percentOfIncorrectAnswers*. Therefore, we can conclude that statistically significant differences exist between the two groups and that on average the Experimental group obtains 2.145 points more in *diffScore* than the Control group. Additionally, every 10% increase in the number of incorrect answers corresponds on average to a *diffScore* decrease of 1.02 points.

When *correctedLearningGain* is considered as the dependent variable, we obtain correctedLearningGain = 0.349 + 0.182 × Group−0.010 × percentOfIncorrectAnswers, with a p-value = 0.040 associated with *Group* and a p-value <0.001 associated with

TABLE 3 Average results (with the standard deviation in parenthesis) obtained in experiment 1A

Variable	1A—Control group	1A—Experimental group
#Students	19	16
preTestScore	−0.158 (1.381)	−0.250 (1.458)
numberOfQuestions	90.000 (56.264)	21.250 (9.713)
numberOfIncorrectAnswers	17.263 (6.181)	6.563 (2.421)
percentOfIncorrectAnswers	26.543 (16.055)	34.879 (16.274)
postTestScore	0.719 (2.337)	1.917 (2.589)
diffScore	0.877 (2.283)	2.167 (3.534)
correctedLearningGain	0.079 (0.227)	0.177 (0.349)

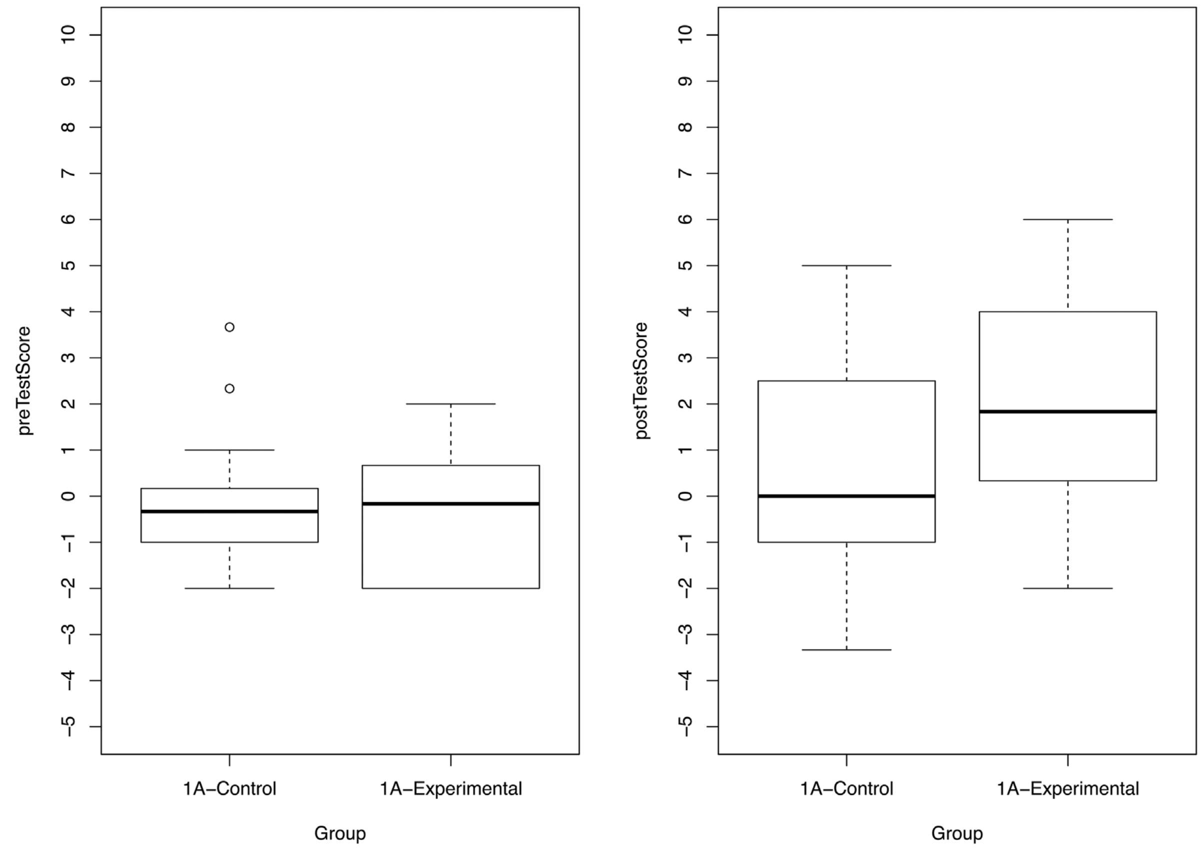

FIGURE 6 Boxplots of the *preTestScore* (left) and *postTestScore* (right) for experiment 1A

percentOfIncorrectAnswers. Therefore, we obtain a similar conclusion as in the previous regression fit.

A total of 35 students answered the anonymous survey. The results for the first four questions in Table 2 are shown for each group in Table 4. From this table, we observe that students evaluated the activity and the testing tool very positively. Additionally, there is a statistically significant difference between both groups in terms of the overall rating of the activity (Student *t*-test, *p*-value <0.001), but not in terms of the testing tool (Student *t*-test, *p*-value = 0.892). Regarding willingness to perform similar activities to those performed in the experiment in the computer laboratory or at home, there are no statistically significant differences between groups (Chi-square test, *p*-value = 0.504 and *p*-value = 0.782 for the computer laboratory or home, respectively).

Regarding the open questions, some of the students reported the following statements:

1. 6 out of 35 (17%) liked the fact that in the training phase the system indicated to them whether they had answered the question incorrectly and which the correct answer was.
2. 3 out of 35 (9%) liked the fact that they could obtain up to one extra point for their final mark in the subject.
3. 10 out of 35 (29%) argued that they had not worked enough on these concepts in previous classes and that their knowledge was not sound enough to solve the exercises proposed.
4. 5 out of 35 (14%) did not find the pretest phase very useful because they did not obtain any feedback from the system.
5. 4 out of 19 (21% of the Control group) would have liked to have received some explanations about their mistakes.
6. 11 out of 16 (69% of the Experimental group) reported that watching mini-videos was what they liked most.
7. 4 out of 16 (25% of the Experimental group) considered the mini-videos very useful for gaining an understanding of the concepts.

4.2 | Experiment 1B

A total of 166 students attended the theoretical lectures in two different groups with the same teacher and 112 out of them participated in experiment 1B: 60 in the Control group and 52 in the Experimental group. The same variables as in the previous experiment were considered and analyzed in the same way. In Table 5, we observe again that both groups have a negative *preTest Score*. The standard Cronbach's alpha was

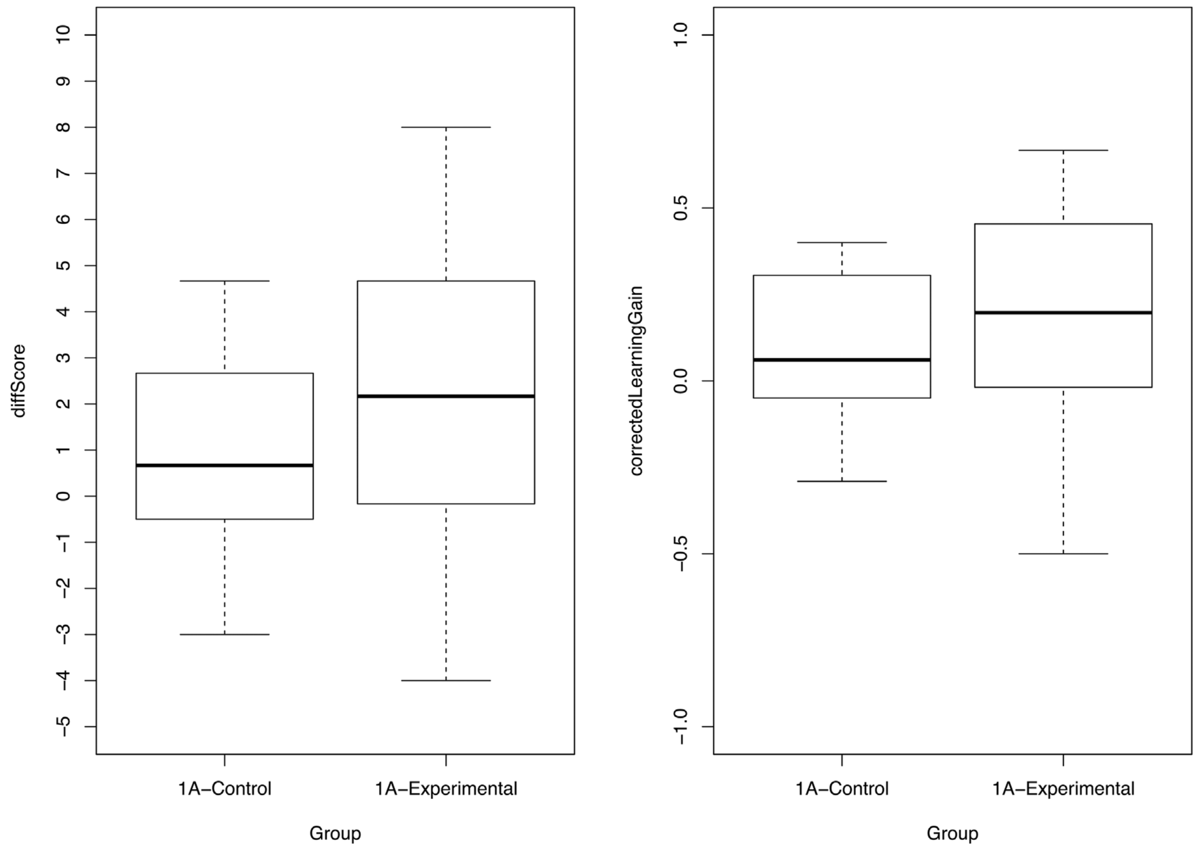

FIGURE 7 Boxplots of the *diffScore* (left) and *correctedLearningGain* (right) for experiment 1A

equal to 0.84 for the *preTest Score* and 0.93 for the *postTest Score*, indicating a highly reliable test for both cases. Figure 8 shows a graphical summary of the *preTest Score* and *postTest Score* variables and Figure 9 shows a graphical summary of the *diffScore* and *correctedLearningGain* variables.

First, we can consider that the baseline knowledge of each group is the same because there is no statistical significance in the *preTest Score* variable (Mann–Whitney–Wilcoxon, *p*-value = 0.801). In other words, we can assume that both groups are homogeneous, indicating that the random assignment was successful in the sense that both groups are comparable before the training session.

Second, regarding the *diffScore* variable there are statistically significant differences in the Experimental group (paired *t*-test, *p*-value <0.001) and in the Control group (Wilcoxon test, *p*-value <0.001). The effect sizes (measured in terms of Cohen's d) were 0.706 for the Experimental group and 0.621 for the Control group, which are considered moderate effects. Regarding the *correctedLearningGain* variable there are statistically significant differences in the Experimental group (paired *t*-test, *p*-value <0.001) and in the Control group (Wilcoxon test, *p*-value <0.001). The effect sizes (measured in terms of Cohen's d) were 0.573 for the Experimental group and 0.664 for the Control group, which are considered moderate effects.

If we now take into account not only the *Group* (0 for the Control group and 1 for the Experimental group) but also the percentage of incorrect answers during the training phase, *percentOfIncorrectAnswers*, we obtain the following linear regression fits.

TABLE 4 Results of the survey administered to the students in experiment 1A

	1A—Control group	1A—Experimental group
Question	**Mean (SD)**	**Mean (SD)**
1. What is your overall rating of the activity? (0–10)	7.00 (1.25)	8.75 (1.13)
2. What is your overall rating of the testing tool? (0–10)	7.32 (1.49)	7.38 (1.02)
	Yes/No	Yes/No
3. Would you like to have similar activities in the lab? (Yes/No)	78.95%/21.05%	87.50%/12.50%
4. Would you like to have similar activities available at home? (Yes/No)	78.95%/ 21.05%	75.00%/25.00%

TABLE 5 Average results (with the standard deviation in parenthesis) obtained in experiment 1B

Variable	1B—Control group	1B—Experimental group
#Students	60	52
preTestScore	−0.375 (1.550)	−0.173 (1.898)
numberOfQuestions	18.533 (18.675)	24.000 (24.297)
numberOfIncorrectAnswers	7.400 (7.883)	7.115 (4.278)
percentOfIncorrectAnswers	48.827 (23.512)	42.024 (18.941)
postTestScore	1.250 (2.673)	1.513 (3.003)
diffScore	1.625 (2.616)	1.686 (2.387)
correctedLearningGain	0.151 (0.264)	0.172 (0.260)

Considering *diffScore* as the dependent variable, we obtain: diffScore = 3.339−0.178 × Group−0.035 × percentOfIncorrectAnswers, with a *p*-value = 0.701 associated with *Group* and a *p*-value = 0.001 associated with *percentOfIncorrectAnswers*, and diffScore = 3.227−0.034 × percentOfIncorrectAnswers, with a *p*-value = 0.001 associated with *percentOfIncorrectAnswers*. Therefore, we can conclude that there are no statistically significant differences between the Experimental and Control groups and that every 10% increase in the number of incorrect answers corresponds in average to a *diffScore* decrease of 0.34 points.

If we now take into account *correctedLearningGain* as the dependent variable, we obtain: correctedLearningGain = 0.360−0.008 × Group−0.004 × percentOfIncorrectAnswers, with a *p*-value = 0.869 associated with *Group* and a *p*-value <0.001 associated with *percentOfIncorrectAnswers*, and correctedLearningGain = 0.355−0.004 × percentOfIncorrectAnswers, with a *p*-value <0.001 associated with *percentOfIncorrectAnswers*. Therefore, we obtain a similar conclusion as in the previous regression fit.

A total of 73 students answered the anonymous survey, 40 out of 60 from the Control group and 33 out of 52 from the Experimental group. Although there are no statistically significant differences between groups (Chi-square test, *p*-value = 0.722), we cannot be certain that the participants in the survey represent an unbiased sample because motivated

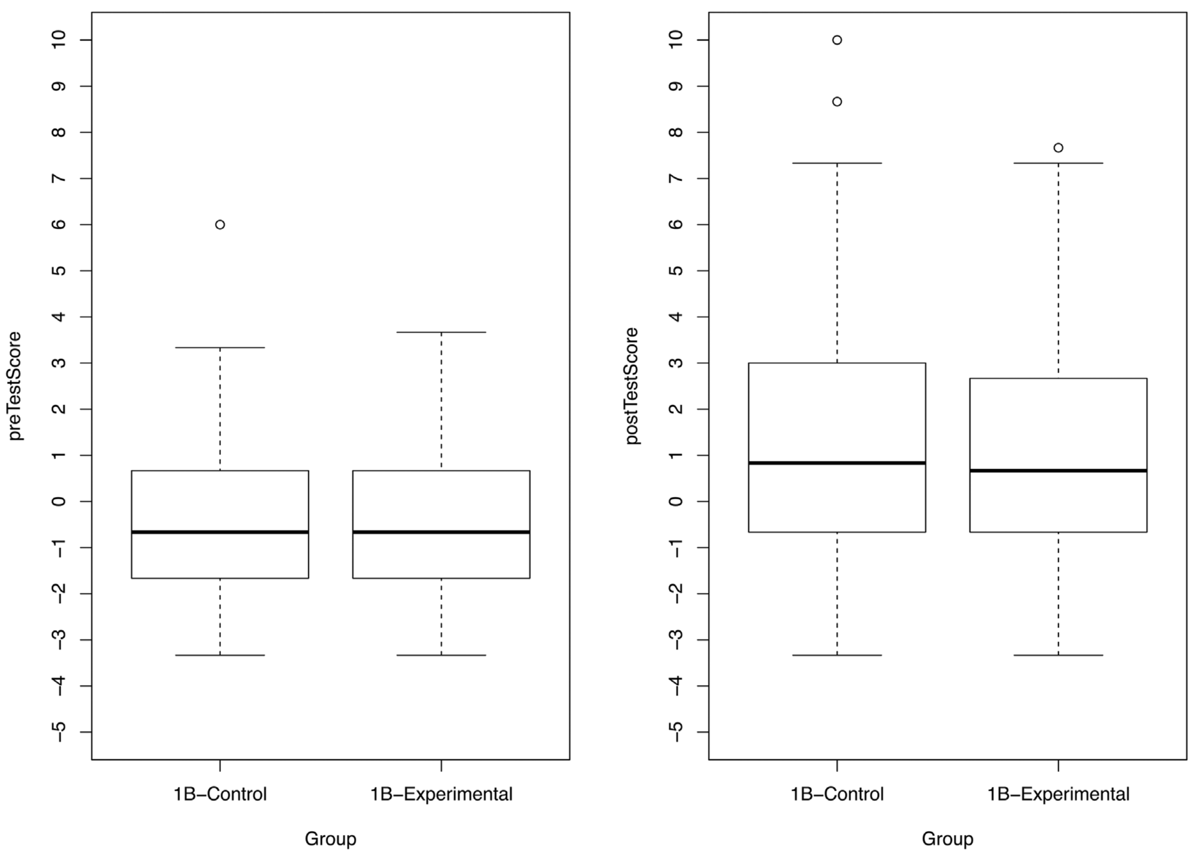

FIGURE 8 Boxplots of the *preTestScore* (left) and *postTestScore* (right) for experiment 1B

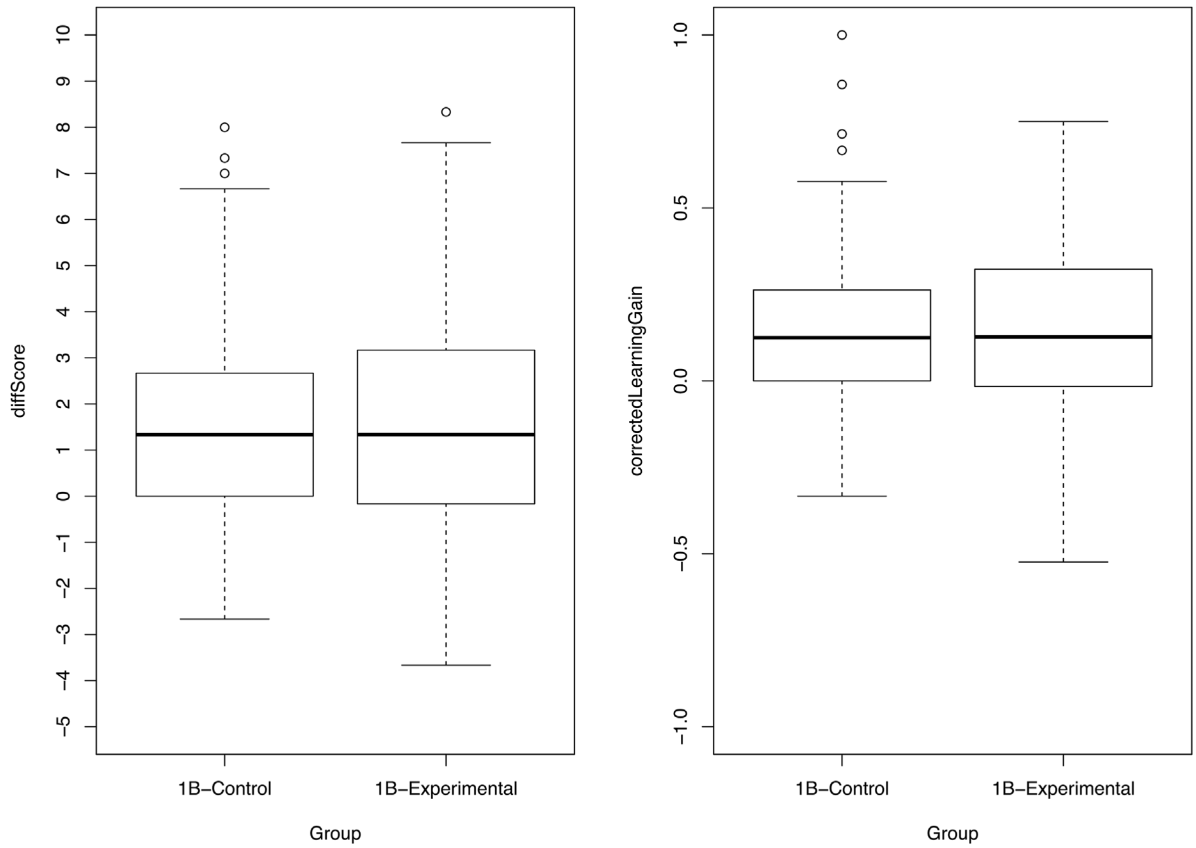

FIGURE 9 Boxplots of the *diffScore* (left) and *correctedLearningGain* (right) for experiment 1B

people or people who liked the activity and/or the testing tool are perhaps more likely to fill out the survey. The results for the first four questions in Table 2 are shown for each group in Table 6. From this table, we observe that students evaluated the activity and the testing tool positively without statistically significant differences between groups (Student *t*-test, p-value = 0.732 and p-value = 0.356, respectively). Regarding willingness to perform similar activities to those performed in the experiment in the computer laboratory or at home, there are no statistically significant differences between groups (Chi-square test, p-value = 0.197 and p-value = 0.519 for the computer laboratory or home, respectively).

Regarding the open questions, some of the students reported the following statements:

1. 12 out of 73 (16%) found that the feedback consisting of an explanation of the concept related to the question is very useful for understanding the subject.
2. 8 out of 73 (11%) liked the fact that in the training phase the system indicated to them whether they had answered the question incorrectly and which the correct answer was.
3. 6 out of 73 (8%) think that this new way of learning is entertaining and helps them to break the routine.
4. 5 out of 73 (7%) liked the fact that they could obtain up to one extra point for their final mark in the subject.
5. 19 out of 73 (26%) stated that they had not done enough work on these concepts in previous classes and that their

TABLE 6 Results of the survey administered to the students in experiment 1B

	1B—Control group	1B—Experimental group
Question	**Mean (SD)**	**Mean (SD)**
1. What is your overall rating of the activity? (0–10)	7.65 (1.73)	7.79 (1.73)
2. What is your overall rating of the testing tool? (0–10)	6.45 (1.69)	6.82 (1.70)
	Yes/No	Yes/No/No answer
3. Would you like to have similar activities in the lab? (Yes/No)	85.00%/15.00%	72.73%/ 24.24%/3.03%
4. Would you like to have similar activities available at home? (Yes/No)	62.50%/37.50%	69.70%/ 30.30%/0.00%

knowledge was not sound enough to solve the exercises proposed.

6. 8 out of 73 (11%) did not find the pretest phase very useful because they did not obtain any feedback from the system.
7. 3 out of 73 (4%) were unhappy about the date of the experiment because the next day they had an exam in another subject.
8. 2 out of 73 (3%) suggested that the training phase should have lasted longer.
9. 2 out of 73 (3%) mentioned that the tests should be available for the students to work at any time in the lab or at home.
10. 7 students out of 40 (18% of the Control group) liked the mini-videos very much.
11. 4 students out of 40 (10% of the Control group) criticized the system for being slow when loading the questions.

5 | DISCUSSION AND CONCLUSIONS

In this paper, we have first presented a hybrid web application to provide video feedback in a formative assessment environment. It has been constructed from two learning resources: Siette, a web-based formative assessment environment, and a set of learning video podcasts known as "modular teaching mini-videos." A key feature that made integration possible was the modularity of this type of video podcasts. The video content was created following the research line proposed by ref. [30] taking into account features that, according to previous works, have demonstrated to be effective for learning.

Our first objective was to investigate whether the combination of formative assessment and video feedback is effective. To prove this hypothesis, we designed a controlled-condition experiment with pre- and posttests to compare the results of a set of students that received video feedback during the training phase (specifically, *KCR + MTM*) with a set of students that received just knowledge of correct response feedback (that is, *KCR*). Based on the regression coefficients used for explaining the student's "learning increase score" (measured in terms of either *diffScore* or *correctedLearningGain*), it is observed a statistically significant positive effect of the type of feedback in favor of *KCR + MTM* and a statistically significant negative effect of the percent of the student's incorrect answers during the training phase (*percentOfIncorrectAnswers*).

Another controlled-condition experiment was designed to compare *KCR + MTM* with *KCR + equivalent illustrated text* in a formative assessment environment. Based on the regression coefficients to explain the student's "learning increase score," it is observed a statistically non-significant effect of the type of feedback and a statistically significant negative effect of *percentOfIncorrectAnswers*. Surprisingly, these results contradict our second hypothesis that video feedback would be superior to equivalent illustrated text feedback. We formulated that hypothesis because the context of our research (Statistics) requires the acquisition of procedural knowledge as in the studies carried out by ref. [32,14], that showed the superiority of the video over the text format. This opposite result can be explained by the fact that neither of these authors considered equivalent illustrated text and is in line with [40] that did not find significant differences between learning from video podcast and equivalent illustrated text in a declarative learning context (German history). As far as we know, our work is the first that compares the effectiveness of equivalent illustrated text and video podcast in a formative assessment environment that requires the acquisition of a procedural learning. Although we have not find statistically significant differences between video podcast and equivalent illustrated text, we believe that it is useful to have both formats following [46] that proposed to vary learning activities as a way of increasing the interest and motivation of the students.

Our two experiments were complemented with two voluntary and anonymous surveys in each group to know the students' perspective and viewpoint. The overall evaluation of the experiments was positive for both the training phase and the testing tool. In the first experiment, it was found a statistically significant difference between *KCR + MTM* and *KCR* in terms of the overall rating of the activity but not in terms of the testing tool. However, in the second experiment, neither the overall rating of the activity nor the testing tool showed a statistically significant difference between *KCR + equivalent illustrated text* and *KCR + MTM*. Students indicated that they would like to perform similar activities in the laboratory and at home. Additionally, all the students evaluated the experiment positively, especially those that received elaborate feedback, who rated it higher. Moreover, the students who did not receive elaborate feedback mentioned that they wished they had received explanations when they had made mistakes in the tests, which can be interpreted as a sign that they wanted to receive elaborate feedback to improve their learning. It is worth mentioning that the fact the students did not report differences in the overall score of the activity in the second study is an indication that the efforts dedicated to obtain equivalent materials have met.

Further research and development will include full personal adaptation of the hybrid web application developed for this study. Currently, the system adapts its behavior to the students' responses, showing the corresponding elaborate feedback only in the event of an incorrect answer. There are three other possible sources for adaptability: (i) adapting the feedback modality (either *KCR + MTM or KCR + equivalent illustrated text*) to the student's preferences; (ii) adapting the feedback depending on the student's response; and (iii) selecting the next question to pose by estimating the student's knowledge using computer adaptive testing [55,45,42]. These features are already part of Siette [21], but developing the appropriate content and collecting enough data to calibrate the student model require considerable effort. We are beginning to work in this direction. As future work, we are also considering the development of similar resources to learn about statistical software tools such as R and MATLAB.

ACKNOWLEDGMENT

The authors would like to thank the Department of Statistics of UC3M for the facilities where the experiments took place.

ORCID

Emilio Letón http://orcid.org/0000-0001-6908-9590
Elisa M. Molanes-López http://orcid.org/0000-0003-3217-8551
Manuel Luque http://orcid.org/0000-0003-3018-3760
Ricardo Conejo http://orcid.org/0000-0003-0810-4608

REFERENCES

1. R. E. Bennett et al., *Using multimedia in large-scale computer-based testing programs*, Comput. Human Behav. **15** (1999), 283–294.
2. J. B. Biggs, *Teaching for quality learning at university: What the student does,* Society for Research into Higher Education & Open University Press, Maidenhead, UK, 2003.
3. P. Black and D. Wiliam, *Assessment and classroom learning*, Assess. Educ. **5** (1998), 7–74.
4. P. Black and D. Wiliam, *Developing the theory of formative assessment*, Educ. Assess. Eval. Account. J. Personnel Eval. Educ. **21** (2009), 5–31.
5. S. Bloxham and P. Boyd, *Developing effective assessment in higher education: A practical guide*. Open University Press, Maidenhead, 2007.
6. B. Brick and J. Holmes, Using screen capture software for student feedback: Towards a methodology using screen capture software for student feedback: Towards a methodology, (2008).
7. T. Buchanan, *The efficacy of a World-Wide Web mediated formative assessment*, J. Comput. Assist. Learn. **16** (2000), 193–200.
8. A. Cann, *Podcasting is dead. Long live video!*, Biosci. Edu. **10** (2007).
9. H. J. Choi and S. D. Johnson, *The effect of context-based video instruction on learning and motivation in online courses*, Am. J. Distance Edu. **19** (2005), 215–227.
10. R. Conejo et al., *Siette: A Web-based tool for adaptive testing*, Int. J. Artif. Intell. Educ. **14** (2004), 29–61.
11. R. Conejo, E. Guzmán, and M. Trella, *The SIETTE automatic assessment environment*, Int. J. Artif. Intell. Educ. **26** (2016), 270–292.
12. D. S. J. Costa, et al., *A web-based formative assessment tool for Masters students: A pilot study*, Comput. Educ. **54** (2010), 1248–1253.
13. A. Crook, et al., *The use of video technology for providing feedback to students: Can it enhance the feedback experience for staff and students*? Comput. Educ. **58** (2012), 386–396.
14. J. V. Dempsey and R. V. Eck, *Modality and placement of a pedagogical adviser in individual interactive learning*, Brit. J. Educ. Technol. **34** (2003), 585–600.
15. F. Donkor, *The comparative instructional effectiveness of print-based and video-based instructional materials for teaching practical skills at a distance*, Int. Rev. Res. Open Distance Learn. **11** (2010), 96–116.
16. A. Furnham, I. Benson, and B. Gunter, *Memory for television commercials as a function of the channel of communication*, Soc. Behav. **2** (1987), 105–112.
17. D. K. Griffin, D. Mitchell, and S. J. Thompson, *Podcasting by synchronising powerpoint and voice: What are the pedagogical benefits*? Comput Educ. **53** (2009), 532–539.
18. M. Guadagnoli, W. Holcomb, and M. Davis, *The efficacy of video feedback for learning the golf swing*, J. Sports Sci. **20** (2002), 615–622.
19. B. Gunter, A. F. Furnham, and G. Gietson, *Memory for the news as a function of the channel of communication*, Hum. Learn. J. Pract. Res. Appl. **3** (1984), 265–271.
20. B. Gunter, A. Furnham, and J. Leese, *Memory for information from a party political broadcast as a function of the channel of communication*, Soc. Behav. **1** (1986), 135–142.
21. E. Guzmán, R. Conejo, and J. L. Pérez De la cruz, *Improving student performance using self-assessment tests*, IEEE Intell. Syst. **22** (2007), 46–52.
22. L. R. Harris, G. T. L. Brown, and J. A. Harnett, *Understanding classroom feedback practices: A study of New Zealand student experiences, perceptions, and emotional responses*, Educ. Assess. Eval. Account. J. Personnel Eval. Educ. **26** (2014), 107–133.
23. T. Hartsell and S. C.-Y. Yuen, *Video streaming in online learning*, AACE J. **14** (2006), 31–43.
24. J. Hattie and H. Timperley, *The power of feedback*, Rev. Educ. Res. **77** (2007), 81–112.
25. Y. He, S. Swenson, and N. Lents, *Online video tutorials increase learning of difficult concepts in an undergraduate analytical Chemistry course*, J. Chem. Educ. **89** (2012), 1128–1132.
26. M. Henderson and M. Phillips, *Video-based feedback on student assessment: Scarily personal*, Aust. J. Educ. Technol. **31** (2015), 51–66.
27. H.-M. Huang, *Toward constructivism for adult learners in online environments*, Brit. J. Educ. Technol. **33** (2002), 27–37.
28. A. Irons, *Enhancing learning through formative assessment and feedback. Key guides for effective teaching in higher education*. Routledge, Abingdon, UK, 2008.
29. S. Jordan, *Student engagement with assessment and feedback: Some lessons from short-answer free-text e-assessment questions*, Comput Educ. **58** (2012), 818–834.
30. R. H. Kay, *Exploring the use of video podcasts in education: A comprehensive review of the literature*, Comput. Human Behav. **28** (2012), 820–831.
31. R. H. Kay and I. Kletskin, *Evaluating the use of problem-based video podcasts to teach mathematics in higher education*, Comput Educ. **59** (2012), 619–627.

32. J. P. Lalley, *Comparison of text and video as forms of feedback during computer assisted learning*, J. Educ. Comput. Res. **18** (1998), 323–338.
33. W. Leahy and J. Sweller, *Cognitive load theory, modality of presentation and the transient information effect*, Appl. Cognitive Psychol. **25** (2011), 943–951.
34. E. Letón and E. M. Molanes-López, Two new concepts in video podcasts—minimalist slides and modular teaching mini-videos, (2014), 292–297.
35. T. Lunt and J. Curran, *Are you listening please? The advantages of electronic audio feedback compared to written feedback*, Assess. Eval. Higher Educ. **35** (2010), 759–769.
36. J. M. Luzón and E. Letón, *Use of animated text to improve the learning of basic mathematics*, Comput Educ. **88** (2015), 119–128.
37. R. E. Mayer, *Learning and instruction*, 2nd ed., Pearson Merrill Prentice Hall, Upper Saddle River, NY, 2008.
38. R. E. Mayer, *Multimedia learning*. 2nd ed., Cambridge University Press, New York, 2009.
39. J. McCarthy, *Evaluating written, audio and video feedback in higher education summative assessment tasks*, Issues Educ. Res. **25** (2015), 153–169.
40. M. Merkt et al., *Learning with videos vs. learning with print: The role of interactive features*, Learn. Instruc. **21** (2011), 687–704.
41. T. Miller, *Formative computer-based assessment in higher education: The effectiveness of feedback in supporting student learning*, Assessment Eval. Higher Educ. **34** (2009), 181–192.
42. Ö. Özyurt, H. Özyurt, and A. Baki, *Design and development of an innovative individualized adaptive and intelligent e-learning system for teaching-learning of probability unit: Details of UZBEMAT*, Expert. Syst. Appl. **40** (2013), 2914–2940.
43. N. Pachler, et al., *Formative e-assessment: Practitioner cases*, Comput. Educ. **54** (2010), 715–721.
44. A. Paivio, *Mental representations: A dual coding approach*, Oxford University Press, New York, 1990.
45. R. Peredo, A. Canales, and A. Menchaca, *Intelligent web-based education system for adaptive learning*, Expert. Syst. Appl. **38** (2011), 14690–14702.
46. P. R. Pintrich, *A motivational science perspective on the role of student motivation in learning and teaching contexts*, J. Educ. Psychol. **95** (2003), 667–686.
47. A. Rae and P. Samuels, *Web-based personalised system of instruction: An effective approach for diverse cohorts with virtual learning environments*? Comput. Educ. **57** (2011), 2423–2431.
48. S. Schwan and R. Riempp, *The cognitive benefits of interactive videos: Learning to tie nautical knots*, Learn. Instruc. **14** (2004), 293–305.
49. V. J. Shute, *Focus on formative feedback*, Rev. Educ. Res. **78** (2008), 153–189.
50. T. Traphagan, J. V. Kucsera, and K. K. Kishi, *Impact of class lecture webcasting on attendance and learning*, Educ. Technol. Res. Dev. **58** (2010), 19–37.
51. W. Turner and J. West, *Assessment for "Digital first language speakers: Online video assessment and feedback in higher education*, Int. J. Teach. Learn. Higher Educ. **25** (2013), 288–296.
52. F. M. van der Kleij, et al., *Effects of feedback in a computer-based assessment for learning*, Comput. Educ. **58** (2012), 263–272.
53. F. M. van der Kleij, R. C. W. Feskens, and T. J. H. M. Eggen, *Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis*, Rev. Educ. Res. **85** (2015), 475–511.
54. T. van Gog, I. Verveer, and L. Verveer, *Learning from video modeling examples: Effects of seeing the human model's face*, Comput. Educ. **72** (2014), 323–327.
55. H. Wainer and R. J. Mislevy, Item response theory, item calibration, and proficiency estimation, H. Wainer, N. J. Dorans, D. Eignor, R. Flaugher, B. F. Green, R. J. Mislevy, L. Steinberg, and D. Thissen (eds.), *Computerized adaptive testing: A primer*, 2nd ed., Lawrence Erlbaum Associates, Hillsdale, NJ, 2000, pp. 61–101.
56. M. Waldrop, *Online learning: Campus 2.0*, Nat. News **495** (2013), 160–163.
57. J. Walma van der Molen and T. van der Voort, *Children's and adults' recall of television and print news in children's and adult news format*, Commun. Res. **27** (2000), 132–160.
58. T.-H. Wang, *Web-based quiz-game-like formative assessment: Development and evaluation*, Comput. Educ. **51** (2008), 1247–1263.
59. C. Wang, D.-Z. Wang, and J.-L. Lin, *ADAM: An adaptive multimedia content description mechanism and its application in web-based learning*, Expert. Syst. Appl. **37** (2010), 8639–8649.
60. J. West and W. Turner, *Enhancing the assessment experience: Improving student perceptions, engagement and understanding using online video feedback*, Innovat. Educ. Teach. Int. **53** (2016), 400–410.
61. M. B. Wieling and W. H. A. Hofman, *The impact of online video lecture recordings and automated feedback on student performance*, Comput. Educ. **54** (2010), 992–998.
62. K. Wilson, et al., *Improving student performance in a first-year geography course: Examining the importance of computer-assisted formative assessment*, Comput. Educ. **57** (2011), 1493–1500.
63. P Yeung, Jr, T Justice, and R. P. Pasic, *Comparison of text versus video for teaching laparoscopic knot tying in the novice surgeon: A randomized, controlled trial*, J. Minim. Invasive Gynecol. **16** (2009), 411–415.
64. D. Zhang, et al., *Can e-learning replace classroom learning*? Commun. ACM. **47** (2004), 75–79.
65. D. Zhang, et al., *Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness*, Inform. Manage. **43** (2006), 15–27.

E. Letón is born in Madrid in 1966. He has a PhD in Mathematics from the Universidad Complutense de Madrid (UCM) in 2002. He is currently a lecturer in the Department of Artificial Intelligence of the Universidad Nacional de Educación a Distancia (UNED), Spain, where he joined in 2009. Previously, he was lecturer in the Department of Statistics in the Universidad Carlos III de Madrid (UC3M) during 5 years. He has also worked for 15 years in Departments of Planning and Statistics in the banking sector and the pharmaceutical industry. His research interests include Survival Analysis, Nonparametric Tests, PLS, Meta-Analysis, Biostatistics and B-Learning. In @emilioleton, his professional tweets

and his personal website with extended information can be found.

E. M. Molanes-López was born in Vigo in 1976. She received her MSc degree in Mathematics in 2000 from the Universidad de Santiago de Compostela (Spain) and the PhD degree in Statistics (with European mention) in 2007 from the Universidade da Coruña (Spain). During her PhD studies, she had the opportunity to visit several foreign universities: Universiteit Hasselt (Belgium), Université Catholique de Louvain (Belgium), and The University of Texas (Houston, USA). Late in 2007, she joined the Statistics Department of Universidad Carlos III de Madrid as a visiting assistant professor. Nowadays, from October 2015, she is an assistant professor at the Biostatistics Unit of the Department of Statistics and Operations Research of Universidad Complutense de Madrid. Her research interests include Nonparametric Statistics, Survival Analysis, Receiver Operating Characteristic (ROC) curves, and Copula functions. She has co-authored eight papers in JCR-indexed journals (seven in Statistics & Probability and one in Engineering, Electrical & Electronic). You can find more information at her personal website: http://elisamariamolanes.wix.com/emolanes.

M. Luque was born in Madrid, Spain, in 1977. He received a Computer Science bachelor's degree from the University of Málaga, Spain, in 2003, and a PhD degree from the Universidad Nacional de Educación a Distancia (UNED), Madrid, in 2009. He is currently an associate professor in the Department of Artificial Intelligence at UNED, where he has worked since 2003. In addition, he is a member of the Research Center for Intelligent Decision-Support Systems (CISIAD) at the same university. His research interests focus on the theory, methods, and applications of probabilistic graphical models, especially in the fields of medicine and education. Personal web page: www.ia.uned.es/~mluque.

R. Conejo received his PhD degree in Ingeniero de Caminos, Canales y Puertos (Civil engineer) from the Technical University of Madrid, in 1995. He is a full professor in the Languages and Computer Science Department of the University of Málaga, Spain, where he has worked since 1986. In addition, he is the Director of the Research and Applications in Artificial Intelligence Group at the same university. He has more than 25 years of experience in the development of learner models and intelligent educational applications. He has published extensively on this field. His research interests currently focus on adaptive testing, student knowledge diagnosis user modeling, and intelligent tutoring systems. He has also worked on fuzzy logic, model-based diagnosis, multiagent systems, and artificial intelligence applied to civil engineering.

How to cite this article: Letón E, Molanes-López EM, Luque M, Conejo R. Video podcast and illustrated text feedback in a web-based formative assessment environment. *Comput Appl Eng Educ*. 2018;26:187–202. https://doi.org/10.1002/cae.21869