Article

Improving the Teaching of Hypothesis Testing Using a Divide-and-Conquer Strategy and Content Exposure Control in a Gamified Environment

David Delgado-Gómez [1,*], **Franks González-Landero** [2], **Carlos Montes-Botella** [1], **Aaron Sujar** [1,3], **Sofia Bayona** [3,4] **and Luca Martino** [5]

[1]   Department of Statistics, Universidad Carlos III, Getafe, 28903 Madrid, Spain;
       carlos.montes@uc3m.es (C.M.-B.); aaronsujar@gmail.com (A.S.)
[2]   Edison Developments, 44002 Teruel, Spain; gonzalezfranks@hotmail.com
[3]   Department of Computer Engineering, Universidad Rey Juan Carlos, 28933 Madrid, Spain;
       sofia.bayona@urjc.es
[4]   Center for Computational Simulation, N/N Montepríncipe Avenue, Boadilla del Monte,
       28660 Madrid, Spain
[5]   Signal Theory and Communication Department, Universidad Rey Juan Carlos, 28933 Madrid, Spain;
       luca.martino@urjc.es
*   Correspondence: ddelgado@est-econ.uc3m.es

check for updates

Abstract: Hypothesis testing has been pointed out as one of the statistical topics in which students present more misconceptions. In this article, an approach based on the divide-and-conquer methodology is proposed to facilitate its learning. The proposed strategy is designed to sequentially explain and evaluate the different concepts involved in hypothesis testing, ensuring that a new concept is not presented until the previous one has been fully assimilated. The proposed approach, which contains several gamification elements (i.e., points or a leader-board), is implemented into an application via a modern game engine. The usefulness of the proposed approach was assessed in an experiment in which 89 first-year students enrolled in the Statistics course within the Industrial Engineering degree participated. Based on the results of a test aimed at evaluating the acquired knowledge, it was observed that students who used the developed application based on the proposed approach obtained statistically significant higher scores than those that attended a traditional class (p-value < 0.001), regardless of whether they used the learning tool before or after the traditional class. In addition, the responses provided by the students who participated in the study to a test of satisfaction showed their high satisfaction with the application and their interest in the promotion of these tools. However, despite the good results, they also considered that these learning tools should be considered as a complement to the master class rather than a replacement.

Keywords: education; learning environments; educational games; engineering students

1. Introduction

Statistics is a core subject in almost all university degrees. As an example, Garfield and Ahlgren pointed out that the University of Minnesota offered 160 statistics courses, taught by 13 different departments [1]. Learning statistics is important as it provides the capacity of conducting logical reasoning and critical thinking, enhancing interpretation and evaluation skills and facilitating dealing with highly abstract concepts [2].

However, many students find learning statistics difficult and unpleasant as it can be appreciated in several studies that reported high levels of statistical anxiety [3–6] or "statisticophobia" [7–9].

Dykeman observed that, compared to students enrolled in general education courses, students in statistics courses had lower levels of self-efficacy and higher levels of anxiety [10]. Ferrandino indicates that the teaching of statistics can be difficult for many reasons that have persisted over time, across disciplines, and across borders [11]. Ben-Zvi and Garfield pointed out different causes as, for example, that some statistical ideas are complex and/or counter-intuitive, that students lack of the required mathematical knowledge, or that they struggle in dealing with the context of the problem [12]. These difficulties can also be noticed in the large number of research articles devoted to explain students' misconceptions related to several statistical concepts [13–16]. These misconceptions appear in almost all the statistical topics, including descriptive statistics [17,18], probability [1,19], or statistical inference [14]. Therefore, developing tools that facilitate the acquisition of statistical concepts is essential. This need increases with the proliferation of online courses on statistics [20–22]. As an example, Reagan points out that several colleges are offering online statistics courses to enable students to complete their academic curriculum [23]. Mills and Raju also highlight the need to learn about how to effectively implement statistical courses [24].

To facilitate the understanding of the different statistical concepts and to correct the misconceptions, several methodologies have been proposed. Two of the most popular approaches are active learning [25] and the flipped classroom [26]. Active learning was initially defined as "anything that involves students in doing things and thinking about the things they are doing" [27]. A more recent and meaningful definition is the one proposed by Felder and Brent [28]. They define active learning as "anything course-related that all students in a class session are called upon to do other than simply watching, listening and taking notes". The use of active learning techniques in statistical courses has raised students' confidence in understanding statistics [29], increased knowledge retention [30], and improved students' performance [31].

In the flipped classroom methodology, knowledge acquisition takes place out of the classroom so that the class time can be devoted to more interactive activities. These can be designed, for instance, to polish the knowledge or to increase students' motivation. The flipped classroom has shown several benefits when it has been used to teach statistics. Wilson's [26] implementation of a flipped classroom in an undergraduate statistics course had a positive impact in students' attitude and performance. Winquist and Carlson found that students who took an introductory statistics course using the flipped classroom approach outperformed the ones that took the same course in a traditional lecture-based approach [32]. They observed that the implementation of a flipped classroom in an undergraduate statistics course had a positive impact in students' attitude and performance. McGee, Stokes, and Nadolsky adapted the flipped classroom methodology to identify students' misconceptions so they can correct them during the master class [33]. The results in Reference [34], when studying mathematics achievement and cognitive engagement, indicate that students in a flipped class significantly outperformed those in a traditional class and those enrolled in the online independent study class. Moreover, flipped learning with gamification promoted students' cognitive engagement better than the other two approaches. Faghihi et al. conclude that students who used the gamified system scored above the median, and their performance was greater than with the alternative method [35], whereas the results reported by Jagušt, Boticki, and Sin suggest that gamification improved student math performance and engagement [36]. Sailer et al. affirm that gamification is not effective per se, but that different game design elements can trigger different motivational outcomes [37]. In particular, their results show that badges, leaderboards, and performance graphs positively affect competence need satisfaction, as well as perceived task meaningfulness.

This article focuses on hypothesis testing which, according to Brewer [38], is probably the most misunderstood and confused of all statistical topics. This assertion can also be appreciated in the review conducted by Castro-Sotos et al. about the different students' misconceptions of statistical inference [14]. Difficulties in distinguishing sample and population [39], confusion between the null and the alternative hypothesis [40], misunderstandings about the significance level [41],

and misinterpretation of the *p*-value [42] are only some of the different misconceptions related to hypothesis testing.

Apart from the above-described general purpose methodologies that have been used to facilitate the learning of any statistical topic, there are some research works devoted to teach hypothesis testing. One of the first works is the one conducted by Loosen, who built a physical device to explain hypothesis testing [43]. He found that his students were enthusiastic about the apparatus, but he did not report any quantitative measure. Schneiter [44] built two applets for teaching hypothesis testing. However, just like Loosen, she did not report about their performance. A more recent work is the one conducted by Wang, Vaughn, and Liu [45]. These researchers showed that the use of interactive animations improves the understanding of hypothesis testing but not students' confidence. Other works have proposed the use of specific examples, such as a two-headed coin [46] or a fake deck that only contains red cards (instead of half deck being red cards and the other half being black cards), to demonstrate hypothesis testing [47].

In this article, a new approach is proposed to facilitate the understanding of hypothesis testing, based mainly on the divide-and-conquer strategy, on the content exposure control, and on the use of a gaming environment that increases student motivation.

2. Material and Methods

The proposed approach is supported by six pillars. These aim at providing a solution to the factors that we assume are responsible for the difficulties in learning hypothesis testing: individual learning differences, misconceptions, knowledge gaps, and lack of attitude and motivation. The six pillars are:

- *Divide and conquer.* Our approach requires an analysis of the subject to be taught, in our case hypothesis testing, to extract which are the main concepts and which are the dependencies between them. In this way, it is possible to plan the acquisition of knowledge in a progressive way, so that students can build the necessary scaffolds to master the subject [48].
- *Flipped classroom methodology.* Not all students assimilate concepts at the same speed. This causes that, in the master class, many students tend to tune out. The fact that, in the proposed approach, the statistical concepts are explained with short videos [49] allows students to watch them as many times as necessary, following their own pace, which facilitates self-directed learning [50]. These videos cover the key concepts of hypothesis testing, as suggested by Pfannkuch and Wild [51].
- *Content exposure control.* In almost all learning platforms, the student can freely move from one concept (video/chapter) to another. In hypothesis testing, this approach might cause major complications because the different concepts are presented sequentially and each is based on the previous ones. If the student has an erroneous concept, its negative effects will extend to the following concepts. For this reason, in the proposed approach, each new concept will not be presented until the student demonstrates sufficient knowledge of the current concept.
- *Formative assessment and feedback.* Our approach evaluates the acquired knowledge as the student learns and provides useful formative feedback to assist the learning process. Whenever students make a mistake, they are informed and appropriate feedback is provided to narrow the student's possible knowledge gap, favoring the learning based on scaffolding [48].
- *Active learning.* The proposed approach adopts a problem-based learning scheme [52]. After each video, the student receives a corresponding set of questions that mimics real situations. These questions are useful for both formative and summative assessment [53]. On the one hand, the feedback can be used by students to improve their learning. On the other hand, these questions also help to assess students' knowledge.
- *Gamification.* It has been shown that the use of game elements increases students' engagement [54]. In a recent review paper, Boyle et al. showed the potential benefits of including games, animations,

and simulations in the teaching of statistics [2]. For those reasons, several game elements are included to increase students' motivation and attitude (e.g., scores, leader-board, or a ranking).

The following section describes how the proposed approach, based on the previous pillars, is implemented in an application. After that, in Section 2.2, the design of an experiment, aiming at evaluating the performance of the proposed approach, is described. The obtained results are presented in Section 3. Finally, the article concludes in Section 4 with a discussion of the obtained results.

2.1. The Hypothesis Testing Learning Tool

The learning tool has been developed using the game engine Unity3D [55] and the Next-Gen User Interface (NGUI) library [56]. Following, its main components are described, beginning with an explanation of the main screen and its elements: buttons that lead to video lectures and problems, and the game elements. After that, the video lectures, problems, and content exposure control mechanisms are explained in more detail.

2.1.1. The Main Screen

Figure 1 shows the main screen of the application, in which its different elements are shown. In particular, it contains the following elements:

Figure 1. Main screen of the application.

- **Buttons**. The application has two types of buttons: circular and square. Each circular button leads to a video in which a key concept of hypothesis testing is explained. The concept explained in each video is indicated on the label placed under each button. The courseware includes five concepts that are: type of contrast (mean, variance, proportion), sample vs. population, the null hypothesis, the alternative hypothesis, and the p-value and its interpretation to obtain conclusions. The square buttons present the student with a series of questions related to the concept explained

in the previous circular button. As it will be detailed, these questions are intended to ensure that the student does not misunderstand any concept. Until the student demonstrates sufficient knowledge of these questions, the following videos and problems are blocked.

- **Stars**. Next to each circular button, there are two stars that can be illuminated or not. If they are not lit, it means that the student does not have enough knowledge about the current concept yet. In this case, the following buttons are deactivated so that the student cannot access to them. When the two stars are illuminated, the student has successfully completed all the problems at this level. If only one of the two stars is illuminated, it means that the student has demonstrated a sufficient level of the current concept, even if he did not correctly completed all the questions asked. When the student gets one or two stars, the following concept is unlocked.
- **Ship's boy and flag bearer**. The icon of a ship's boy indicates to students their progress in the course. In this way, it is easy to appreciate the concepts that are already mastered and those that still need to be completed. In addition, there is an icon of a flag bearer pirate that indicates the position of the most advanced student in the session. This second icon allows students to know their position with respect to the leader student of his session.
- **Score and leader-board**. Students receive points as they complete the various problems according to a scoring system that will be explained later. In the lower part of the screen, the achieved points can be seen, whereas, in the upper right corner, there is a leader-board that shows the classification of the top 10 students of the session based on their scores.

2.1.2. Videolectures and Questions

As mentioned, the buttons lead to video lectures and assessment questions. Next, we describe their functioning.

Videolectures

Each video explains one of the five above discussed concepts and has an approximate average duration of 6 min. Students can watch each video as many times as desired so that if they are not able to solve the problems related to a concept, they can watch the associated video again. The only restriction is that students can only watch the videos that they have unlocked.

Problems and Questions

As mentioned, our approach is based on a problem-based learning strategy [52]. For the construction of the problems, the difficulty level has been designed to stay within the student's Zone of Proximal Development (ZPD) [57], keeping it understandable and reachable, but challenging enough.

Each of the five key concepts is evaluated with six questions. These questions are obtained from six practical problems that the student must solve. One of these problems is shown in Figure 2. It can be observed that, in this problem, the student must identify the type of contrast, recognize the data sample, formulate the null hypothesis, enunciate the alternative hypothesis, and, based on the *p*-value, determine if it is possible to reject the null hypothesis or not. It is important to note that students do not initially receive all the questions related to a problem at once. Instead, they only receive the question, for each problem, corresponding to the current concept, remaining within their ZPD. Questions related to more advanced concepts appear invisible until the student has reached that level. That is, if a student has currently watched the video that explains the concept of the data sample, he/she will only receive the second question for each of the six problems. Below, these questions are described in more detail.

1. **Type of contrast**. The first question evaluates whether the student is able to recognize each one of the three types of contrasts that are explained in the first video (mean, variance and proportion). For this, the student selects the type of contrast that he considers associated with the problem statement from a drop-down window.

2. **Data sample**. The second question evaluates whether the student is able to differentiate between sample and population, both with regards to the concepts and to their notation. To do this, the student must identify the problem values belonging to the sample and enter them in the correct text boxes, that is, in the boxes labeled with the corresponding notation to the sample statistics and the sample size. In addition to these text boxes, there are other text boxes related to different population parameters in which the student must not include any value.

3. **Null hypothesis H_0**. The objective of the questions associated with this third concept is to determine if the student is able to recognize the null hypothesis and to write it correctly, using the appropriate notation of the population and not using the one in the sample. For this, the student must recognize the data of the problem that corresponds to the null hypothesis and place it in the corresponding box.

4. **Alternative hypothesis H_1**. The fourth question evaluates if the student is able to identify the alternative hypothesis H_1.

5. **p-value and conclusion**. Finally, students must determine according to the answers given to the previous questions, and to the p-value that is provided, if the null hypothesis can be rejected or not at a significance level α of 0.05. To do this, they must choose the option that they consider correct between the two possibilities presented to them.

Figure 2. One of the problems (translated to English) included in the courseware.

As previously mentioned, these questions evaluate students' knowledge about each concept and are responsible for determining if a student can advance to the next concept or not. When a student correctly answers the six questions, he/she receives two stars, and, when he answers four or five questions correctly out of the six, the student receives one star. In these two cases, the following concept is unlocked, and the student can move forward. When the student answers less than four questions correctly, he/she does not receive any star and must redo these questions. Following the scaffolding approach, each time the student answers a question erroneously, he/she receives formative feedback in the way of a message that helps him/her to identify his/her error, facilitating his progress. Therefore, it is important to emphasize that the questions do not only assess students' knowledge; the fact of receiving feedback and having to repeat the questions until achieving a minimum knowledge allows students to better understand the concept and to correct their misconceptions.

In addition, the student receives 10 points per question answered correctly on the first attempt. As there are 6 questions, it is possible to obtain a maximum of 60 points per level. The questions answered correctly on the second and subsequent attempts provide 5 points. The student must repeat

the level if he/she was not able to correctly answer up to four questions. Students that correctly answered four or five questions can repeat a level, until having correctly answered the six questions and obtaining the two stars.

2.2. Evaluation of the Approach Application

An experiment was conducted to evaluate the benefits of applying the proposed approach with respect to a traditional master class. Following, the details of the experiment are presented.

2.2.1. Sample

A total of 89 students, who were studying the subject of Statistics belonging to the degree of Industrial Engineering, participated in the study. Students were informed that their participation was voluntary and was under no circumstances considered in their academic evaluation. In order to make the data collected anonymous, a random number was given to each student to access the application and sign the tests with it. None of the researchers involved in data collection and analysis were teaching the students who participated in the experiment. After the study, all students were granted access to the application and all the material.

2.2.2. Assessment of Learning Methods

In order to assess the students' knowledge acquisition, two comparable problems were prepared. These two problems were intended to determine whether the student had been able to identify: (i) the type of contrast; (ii) the data pertaining to the sample together with the appropriate notation; (iii) the null hypothesis; (iv) the alternative hypothesis; and, finally, (v) the interpretation of the p-value in relation to the problem in question. Each of the problems was scored either with 1 (totally correct), 0.5 (partially correct), or 0 (incorrect). These two problems were:

Question (Q1): The average expenditure per customer in a store was 89 euros before the recession. Currently, taking a sample of 70 shopping carts, an average of 86 euros with a standard deviation of 9 euros is obtained. According to these data, and assuming a significance level of 0.05, could we affirm that you can see the effect of the current recession? Provide your answer by identifying the type of contrast, the sample data, the null and alternative hypothesis, assuming a p-value of 0.003.

Question (Q2): The ideal weight for 1.80 m tall men is 75 kg. Given a sample of 45 men that are 1.80 m tall in Spain, the average weight turns out to be 77 kg with a standard deviation of 8.5. According to these data, can we say that the Spanish are too fat? Provide your answer by identifying the type of contrast, the sample data, the null and alternative hypothesis, assuming a p-value of 0.06 and a significance level of 0.05.

2.2.3. Experimental Set-Up and Study Groups

The 89 students who participated in the experiment were divided in 3 groups according to their interest in using the application and their time availability. Students who decided to use the application indicated their time availability to participate before and/or after the master class. With this information, we tried to balance the number of participants in each of the groups as much as possible. The first subgroup, which will be referred as G1 from hereafter, was composed of 10 students. The students belonging to this group used the developed application some days before the class during a session of one hour. These students did not have any prior knowledge about hypothesis testing when they participated in this activity. Days later, they attended to the master class in which hypothesis testing was explained. The second group, which will be referred as G2, was the control group. It was made up of 60 students who attended only to the master class. Finally, the group G3 contained the remaining 19 students who enrolled for attending to the master class and days later had access to the application for one hour. There were two students of the G1 who used the application, but they did not attend later to the master class. There were also 4 students from group G3 who attended to the master class, but they did not appear later to use the application.

Next, we describe the order in which each group answered the assessment questions. Group G1 received the Q1 question immediately after their session (prior to the lecture) and the Q2 question after the lecture. Groups G2 and G3 received the Q1 question after the lecture class (none of them had performed used the application). In this way, it was possible, on the one hand, to use Q1 to compare the application of the proposed approach and the master class. On the other hand, it also allowed comparing the performance of groups G2 and G3, which is important to ensure that there are no differences between students who had signed up for using the application from those who did not, as both groups had simply attended the master class by the time they answered the question Q1. Lastly, students belonging to G3 answered question Q2 after using the application. In this way, it was possible to analyze if it is preferable that students to conduct the session with the application before or after the lecture. The design of the experiment is shown in Figure 3. From now on, we will refer to the different sets of scores through the pair formed by the student group and the question presented.
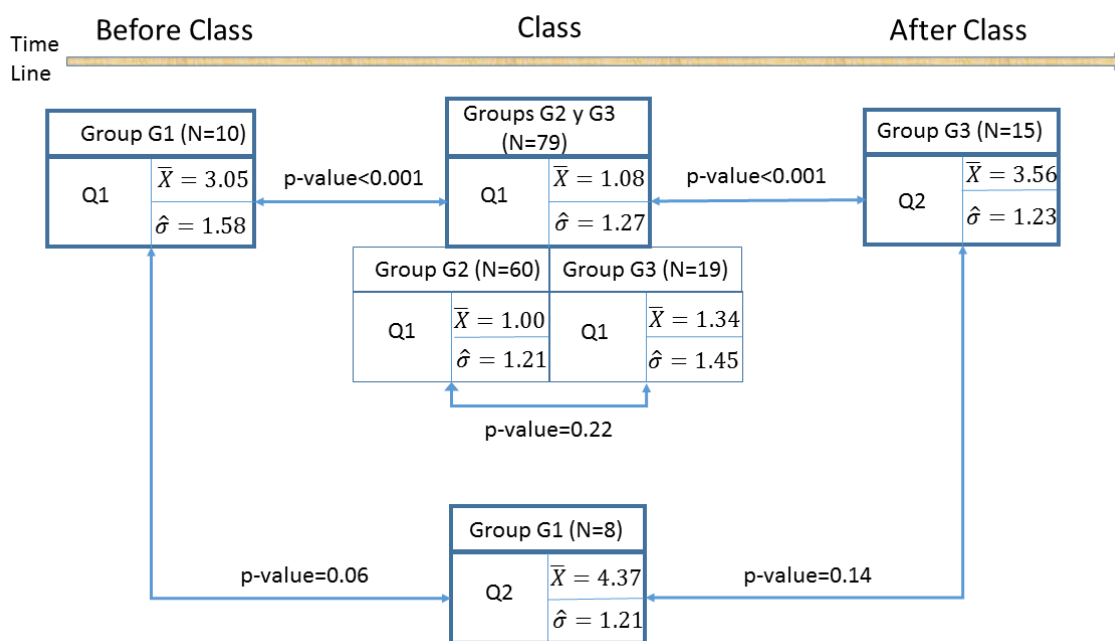

Figure 3. Experimental design including the 3 study groups (G1, G2, and G3), the two assessment questions (Q1 and Q2), and the results of the statistical analysis.

Additionally, the students who participated in the extra session filled out a questionnaire to evaluate the acceptance of the application. This questionnaire contained eight five-point Likert items and two open-ended questions.

3. Results

In this section, we present the results obtained. Figure 3 shows the mean and the standard deviation of the scores obtained for each one of the participant groups in the different phases of the study. Before examining these data and the significance of the differences, it is important to present a result obtained when analyzing the distribution of the scores. This analysis revealed that the distribution of the scores obtained by the students who used the application followed a Gaussian distribution, whereas, when students had only attended to the lecture class, it followed an exponential distribution. Concisely, the goodness-of-fit test to normality Kolmogorov–Smirnov yielded p-values of 0.99, 0.38, and 0.51 for the answers obtained in the pairs-question group G1-Q1, G1-Q2, and G3-Q2, respectively. Similarly, a Chi-square goodness-of-fit test showed that the distribution scores in G2-Q1 and G3-Q1 followed an exponential distribution (p-values 0.14 and 0.18, respectively).

After analyzing the distribution of scores, several hypothesis tests were performed to understand the results obtained. First, we compared the students' scores in the scenario (G1, Q1) with those obtained in (G2 and G3, Q1). A *t*-test showed that the differences were significant (*p*-value < 0.001). This result indicates that the proposed approach seems to be more effective to teach hypothesis test concepts than the lecture class.

Similarly, a *t*-test also showed that the differences between (G2 and G3, Q1) and (G3, Q2) were also significant (*p*-value < 0.001). This result also indicates that, similarly to the previous result, the application of the proposed approach after the master class significantly improves the knowledge acquired by students during the lecture.

Another interesting comparison was the one between the scores obtained in (G1, Q1) and (G1, Q2). The corresponding *t*-test provided a *p*-value close to 0.05, which seems to indicate that the lecture class reinforces the knowledge obtained initially with our approach.

In the comparison of the scores (G1, Q2) and (G3, Q2), the *t*-test indicated that the results were not significant (*p*-value 0.14). However, although this difference is not significant, performing the application prior to the class seems to provide better results, which is consistent with the flipped classroom methodology approach.

Finally, the scores obtained in (G2, Q1) and (G3, Q1) were compared to determine if the participants who used the application after the master class and those who participated only in the master class were similar. Since the scores in these groups followed an exponential distribution, the non-parametric Mann–Whitney test was used to analyze the difference. This test showed a *p*-value of 0.22, indicating that there were no significant differences between these two groups.

Regarding the subjective acceptance questionnaire about the application of the proposed approach, Table 1 shows the obtained results. This table shows the eight items that were presented to the students together with the mean and the standard deviation of their responses. These numbers indicate that the students considered that the application helps to improve assimilation of concepts and that it should be promoted. This can be seen in the values obtained for items 1 and 2, in which the means were 4.30 and 4.80, respectively. The students also considered that the experience was positive (item 8). However, despite these numbers, the students consider this application as a complement to the course but not as a replacement to the teacher (items 5, 6, and 7). On the other hand, despite that explanatory messages were clarifying (item 3), this questionnaire also allowed discovering elements that need to be improved, such as its comfort of use (item 4). In particular, the participants complained about the fact that our video player did not have the option of reproducing only certain parts of the video.

Table 1. Results from questionnaire about the use of the application.

SATISFACTION QUESTIONNAIRE
Date:
User Id:
Please, answer the following questions related to the application with a number from 1 to 5, using the following codes:
1-strongly disagree, 2-disagree, 3-neither agree nor disagree, 4-agree, 5-strongly agree

Item	Mean Score (Std)
1. I think that this type of application helps to improve the knowledge assimilation of the course	4.30 (0.61)
2. This kind of applications should be promoted	4.80 (0.40)
3. The explanatory messages have been sufficiently clarifying and concise	4.26 (0.72)
4. The application is comfortable to use	3.69 (0.73)
5. This application should not be mandatory	3.24 (1.23)
6. This type of programs is good complementary material for the course	4.53 (0.50)
7. This type of material should replace the tutor / professor	2.15 (0.73)
8. I would rate the experience positively	4.30 (0.47)

In addition to the eight items above, students answered two open-ended questions. The first question asked the student how he/she had felt when his/her name moved up on the leader-board. This question was answered by 14 students, 13 of whom (92.8%) provided answers related to increased motivation, while the remaining student replied that he did not paid attention to the leader-board.

The second question asked students if they had felt the need to overtake other participants on the leader-board. Of the 16 students who answered this question, 13 (81.2%) answered affirmatively and justified their response on the basis of competitiveness. The remaining three students answered that their goal was either to learn or their personal improvement.

The students' subjective satisfaction, together with the objective values shown above assessing students' knowledge, show the benefits of applying the proposed approach.

4. Discussion and Conclusions

In this article, a new approach has been proposed for the teaching of hypothesis testing. It has taken into account different elements that have been proposed in the literature to facilitate statistical learning, such as the inclusion of video lectures that allow students to learn at their own pace and a problem-based learning approach to encourage active learning. Additionally, the proposed approach includes a mechanism for content exposure control. In this way, students cannot access to any new concept until they demonstrate having understood the previously presented concepts. This mechanism provides several benefits. First, it avoids that, if the student has understood a concept in a wrong way, this error spreads to the following concepts. Second, when students make an error, they receive immediate formative feedback about it and have the opportunity to correct errors as they occur, providing a good basis to understand the next concept. In addition, it increases the students' confidence making them capable of solving the different questions. Finally, the application that incorporates the proposed approach contains elements of gamification to increase student's motivation.

In the experiment, carried out to identify the advantages and disadvantages of the proposed approach with respect to a traditional class, it was observed that students assimilated the concepts much better when they benefited from the proposed approach. This was true regardless of whether our approach had been done before or after the traditional class, although results seem to indicate that the advantage is greater if it is performed earlier, which is consistent with the flipped-class learning approach.

The students who used the application based on our approach were able to complete almost perfectly the six questions that were formulated with respect to the five concepts involved in the resolution of a hypothesis test. Conversely, in the traditional class, they correctly answered 1.34 concepts on average, in the best of cases. A possible explanation of this remarkable difference can be attributed to the fact that, in a master class, the different concepts are explained sequentially. If students are not able to fully understand one of them, they will have serious difficulties in understanding the subsequent ones, limiting their participation to taking notes that they will use later to try to understand the subject. Regarding the gamification elements, more than 50% of the students expressed that these elements enhanced their competitiveness and made learning more entertaining.

If we compare our proposal with others in the literature, we observe that our students also experimented difficulties understanding the statistical concepts when following the master class, corroborating the conclusions of the review made by Castro-Sotos et al. [14]. In addition, as suggested in Boyle et al. [2] and Sailer et al. [37], we included game design elements, such as stars as game badges, a leader board, a scoring system to indicate performance, and even a flag carrier to show the progress of the most advanced student in the session. Regarding performance, though most studies on learning statistics propose improvements, few carry out a formal evaluation. Wang, Vaughn, and Li [45] did evaluate performance after using different animation interactivity techniques, and their results indicated that the increase of animation interactivity could enhance student achievement improvement on understanding and lower-level applying but not on student remembering and higher-level applying. In terms of confidence improvements, there were no significant differences between their four groups. In turn, our results, depicted in Figure 3, reflect high student satisfaction and a mean significant difference of about 2 points out of 5 (p-value < 0.001). As discussed above, our work stands out for dividing complexity into simpler concepts that students can only access in a controlled, progressive

manner, and for including elements of gamification, demonstrating a beneficial effect on understanding complex issues and obtaining positive results in terms of student satisfaction.

Some limitations of this research should be noted. First, the generalization of the conclusions of this study to other research contexts, subjects, and group sizes should be studied, due to our limited sample size. Despite this fact, our results showed significant improvement in performance. Another current limitation is that most learning platforms do not include facilities to implement our proposal to control content exposure, which means that it must be specifically implemented for each case. In this study, we created a small learning engine. We hope that, as more studies demonstrate the usefulness of this approach, tools, such as Moodle or EdX, could consider implementing such functionality.

These results show the need to rethink traditional lectures to include the benefits derived from methodologies that incorporate current technologies. The acquisition and understanding of difficult concepts by dividing them into steps or simpler concepts allows students to build the learning. The content exposure strategy not only controls the progress of students, but it is also a very interesting tool for the teacher to see what concepts are hard to master, as well as to evaluate whether it is worth dividing them even further or generating additional material. In addition, it offers the advantage of providing an overview of the overall progress of the group. Our work shows that, just by using the application that implements the methodology, the students' results are reasonably good. This could offer the possibility of redesigning the master class in a different way, to be more oriented to clarify concepts or to increase the cognitive level of learning.

We plan to apply the proposed approach for learning different subjects. In addition, not only gamification elements could be included, but also video games themselves. For example, we are currently developing a video game to teach quality control. In this video game, the student is responsible for controlling the filling of a bottling company. To do this, using an avatar, the student must collect various samples, measure them, and create the control charts. And later, he/she will have to detect whether or not the process is under control.

This work is only a first step, but its extensions are immediate. For example, since the application has been well accepted by the students and its usefulness has been demonstrated even without the intervention of the teacher, it could be offered as a self-learning tool aimed at pre-balancing basic statistical knowledge among students before starting other courses. It would also be interesting to dynamically create specific itineraries for each student, tailored according to the particular results obtained. Future versions can include the adaptive selection of exercises that are based on the detected errors, fostering deliberative practice [58], since this would enable students to pay more attention to their weaker areas. Furthermore, with the growth of online courses, especially with COVID-19, providing learner-centered tools that ensure good understanding becomes even more relevant. We want to finish this article with a phrase that we heard a student say to another when the proposed activity ended: "Now, finally, I understood it all".

Abbreviations

The following abbreviations are used in this manuscript:

ZPD   Zone of Proximal Development

References

1. Garfield, J.; Ahlgren, A. Difficulties in learning basic concepts in probability and statistics: Implications for research. *J. Res. Math. Educ.* **1988**, 44–63. [CrossRef]
2. Boyle, E.A.; MacArthur, E.W.; Connolly, T.M.; Hainey, T.; Manea, M.; Kärki, A.; Van Rosmalen, P. A narrative literature review of games, animations and simulations to teach research methods and statistics. *Comput. Educ.* **2014**, *74*, 1–14. [CrossRef]
3. Onwuegbuzie, A.J.; Seaman, M.A. The Effect of Time Constraints and Statistics Test Anxiety on Test Performance in a Statistics Course. *J. Exp. Educ.* **1995**, *63*, 115–124. [CrossRef]
4. Onwuegbuzie, A.J. Writing a research proposal: The role of library anxiety, statistics anxiety, and composition anxiety. *Libr. Inf. Sci. Res.* **1997**, *19*, 5–33. [CrossRef]
5. Pan, W.; Tang, M. Examining the effectiveness of innovative instructional methods on reducing statistics anxiety for graduate students in the social sciences. *J. Instr. Psychol.* **2004**, *31*, 149–159.
6. Schacht, S.; Stewart, B.J. What's Funny about Statistics? A Technique for Reducing Student Anxiety. *Teach. Sociol.* **1990**, *18*, 52–56. [CrossRef]
7. Dillon, K. Statisticophobia. *Teach. Psychol.* **1982**, *9*, 117. [CrossRef]
8. Onwuegbuzie, A.J. Academic procrastination and statistics anxiety. *Assess. Eval. High. Educ.* **2004**, *29*, 3–19. [CrossRef]
9. Stickels, J.; Dobbs, R. Helping Alleviate Statistical Anxiety with Computer Aided Statistics Classes. *J. Scholarsh. Teach. Learn.* **2012**, *7*, 1–15.
10. Dykeman, B.F. Statistics Anxiety: Antecedents and Instructional Interventions. *Education* **2011**, *132*, 441.
11. Ferrandino, J. Student Achievement in Undergraduate Statistics: The Value of Encouraging Failure. *J. Scholarsh. Teach. Learn.* **2016**, *16*, 1. [CrossRef]
12. Ben-Zvi, D.; Garfield, J. Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*; Springer: Dordrecht, The Netherlands, 2004; Chapter 1, pp. 3–15.
13. Batanero, C.; Godino, J.D.; Vallecillos, A.; Green, D.; Holmes, P. Errors and difficulties in understanding elementary statistical concepts. *Int. J. Math. Educ. Sci. Technol.* **1994**, *25*, 527–547. [CrossRef]
14. Castro-Sotos, A.E.; Vanhoof, S.; Van den Noortgate, W.; Onghena, P. Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educ. Res. Rev.* **2007**, *2*, 98–113. [CrossRef]
15. Koparan, T. Difficulties in learning and teaching statistics: Teacher views. *Int. J. Math. Educ. Sci. Technol.* **2015**, *46*, 94–104. [CrossRef]
16. Garfield, J.; Ben-Zvi, D. How students learn statistics revisited: A current review of research on teaching and learning statistics. *Int. Stat. Rev.* **2007**, *75*, 372–396. [CrossRef]
17. Zaidan, A.; Ismail, Z.; Yusof, Y.M.; Kashefi, H. Misconceptions in descriptive statistics among postgraduates in social sciences. *Procedia-Soc. Behav. Sci.* **2012**, *46*, 3535–3540. [CrossRef]
18. Delmas, R.; Liu, Y. Exploring students' conceptions of the standard deviation. *Stat. Educ. Res. J.* **2005**, *4*, 55–82.
19. Delmas, G.; Joan, G.; Ooms, A.; Chance, B. Assessing students' conceptual understanding after a first course in statistics. *Stat. Educ. Res. J.* **2007**, *6*, 28–58.
20. Larreamendy-Joerns, J.; Leinhardt, G.; Corredor, J. Six Online Statistics Courses: Examination and Review. *Am. Stat.* **2005**, *59*, 240–251. [CrossRef]
21. Allen, I.E.; Seaman, J. *Learning on Demand: Online Education in the United States, 2009*; Sloan Consortium: Newburyport, MA, USA, 2010.
22. Herbert, M. Staying the course: A study in online student satisfaction and retention. *Online J. Distance Learn. Adm.* **2006**, *9*, 300–317.
23. Reagan, A. Teaching Undergraduate Social Statistics Online: A Classroom Assessment of Instructional Interventions to Reduce Statistics Anxiety and Facilitate Student Success. *J. Educ. Soc. Policy* **2018**, *5*. [CrossRef]
24. Mills, J.D.; Raju, D. Teaching Statistics Online: A Decade's Review of the Literature About What Works. 2011. Available online: https://www.tandfonline.com/doi/abs/10.1080/10691898.2011.11889613 (accessed on 29 August 2017).

25. Dolinsky, B. An active learning approach to teaching statistics. *Teach. Psychol.* **2001**, *28*, 55–56.
26. Wilson, S.G. The flipped class: A method to address the challenges of an undergraduate statistics course. *Teach. Psychol.* **2013**, *40*, 193–199. [CrossRef]
27. Bonwell, C.C.; Eison, J.A. *Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports*; The George Washington University: Washington, D.C., USA, 1991.
28. Felder, R.M.; Brent, R. Active learning: An introduction. *ASQ High. Educ. Brief* **2009**, *2*, 1–5.
29. Carlson, K.A.; Winquist, J.R. Evaluating an active learning approach to teaching introductory statistics: A classroom workbook approach. *J. Stat. Educ.* **2011**, *19*, 1–23. [CrossRef]
30. Kvam, P.H. The effect of active learning methods on student retention in engineering statistics. *Am. Stat.* **2000**, *54*, 136–140.
31. Freeman, S.; Eddy, S.L.; McDonough, M.; Smith, M.K.; Okoroafor, N.; Jordt, H.; Wenderoth, M.P. Active learning increases student performance in science, engineering, and mathematics. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 8410–8415. [CrossRef]
32. Winquist, J.R.; Carlson, K.A. Flipped statistics class results: Better performance than lecture over one year later. *J. Stat. Educ.* **2014**, *22*, 1–10. [CrossRef]
33. McGee, M.; Stokes, L.; Nadolsky, P. Just-in-Time Teaching in Statistics Classrooms. *J. Stat. Educ.* **2016**, *24*, 16–26. [CrossRef]
34. Lo, C.K.; Hew, K. A comparison of flipped learning with gamification, traditional learning, and online independent study: The effects on students' mathematics achievement and cognitive engagement. *Interact. Learn. Environ.* **2020**, *28*, 464–481. [CrossRef]
35. Faghihi, U.; Brautigam, A.; Jorgenson, K.; Martin, D.; Brown, A.; Measures, E.; Maldonado-Bouchard, S. How gamification applies for educational purpose specially with college algebra. *Procedia Comput. Sci.* **2014**, *41*, 182–187. [CrossRef]
36. Jagušt, T.; Boticki, I.; So, H.J. Examining competitive, collaborative and adaptive gamification in young learners' math learning. *Comput. Educ.* **2018**, *125*, 444–457. [CrossRef]
37. Sailer, M.; Hense, J.U.; Mayr, S.K.; Mandl, H. How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Comput. Hum. Behav.* **2017**, *69*, 371–380. [CrossRef]
38. Brewer, J.K. Behavioral statistics textbooks: Source of myths and misconceptions? *J. Educ. Behav. Stat.* **1985**, *10*, 252–268. [CrossRef]
39. Chance, B.; del Mas, R.; Garfield, J. Reasoning about sampling distribitions. In *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 295–323.
40. Vallecillos, A. Some empirical evidence on learning difficulties about testing hypotheses. *Bull. Int. Stat. Inst. Proc. Fifty-Second Sess. Int. Stat. Inst.* **1999**, *58*, 201–204.
41. Haller, H.; Krauss, S. Misinterpretations of significance: A problem students share with their teachers. *Methods Psychol. Res.* **2002**, *7*, 1–20.
42. Gliner, J.A.; Leech, N.L.; Morgan, G.A. Problems with null hypothesis significance testing (NHST): What do the textbooks say? *J. Exp. Educ.* **2002**, *71*, 83–92. [CrossRef]
43. Loosen, F. A concrete strategy for teaching hypothesis testing. *Am. Stat.* **1997**, *51*, 158–163.
44. Schneiter, K. Two Applets for Teaching Hypothesis Testing. 2008. Available online: https://www.tandfonline.com/doi/full/10.1080/10691898.2008.11889575 (accessed on 29 August 2017).
45. Wang, P.Y.; Vaughn, B.K.; Liu, M. The Impact of Animation Interactivity on Novices' Learning of Introductory Statistics. *Comput. Educ.* **2011**, *56*, 300–311. [CrossRef]
46. Holland, B.K. A classroom demonstration of hypothesis testing. *Teach. Stat.* **2007**, *29*, 71–73. [CrossRef]
47. Eckert, S. Teaching hypothesis testing with playing cards: A demonstration. *J. Stat. Educ.* **1994**, *2*, 1–4.
48. Belland, B.R.; Glazewski, K.D.; Richardson, J.C. A scaffolding framework to support the construction of evidence-based arguments among middle school students. *Educ. Technol. Res. Dev.* **2008**, *56*, 401–422. [CrossRef]
49. Hund, L.; Getrich, C. A Pilot Study of Short Computing Video Tutorials in a Graduate Public Health Biostatistics Course. *J. Stat. Educ.* **2015**, *23*, 1–15. [CrossRef]
50. Long, H.B. *Self-Directed Learning: Emerging Theory & Practice*; Oklahoma Research Ctr: Oklahoma City, OK, USA, 1989.

51. Pfannkuch, M.; Wild, C. Towards an understanding of statistical thinking. In *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 17–46.

52. Tishkovskaya, S.; Lancaster, G.A. Teaching strategies to promote statistical literacy: Review and implementation. In Proceedings of the 8th International Conference on Teaching Statistics, Ljubljana, Slovenia, 11–16 July 2010; pp. 1–4.

53. Harlen, W.; James, M. Assessment and learning: Differences and relationships between formative and summative assessment. *Assess. Educ. Princ. Policy Pract.* **1997**, *4*, 365–379. [CrossRef]

54. Muntean, C.I. Raising engagement in e-learning through gamification. In Proceedings of the 6th International Conference on Virtual Learning ICVL, Cluj-Napoca, Romania, 28–29 October 2011; pp. 323–329.

55. Xie, J. Research on key technologies base Unity3D game engine. In Proceedings of the 2012 7th International Conference on Computer Science & Education (ICCSE), Melbourne, VIC, Australia, 14–17 July 2012; pp. 695–699.

56. Bernardoff, C. *NGUI for Unity*; Packt Publ.: Birmingham, UK, 2014.

57. Vygotsky, L.S. *Mind in Society: The Development of Higher Psychological Processes*; Harvard University Press: Cambridge, MA, USA, 1980.

58. Ericsson, K.A. Discovering deliberate practice activities that overcome plateaus and limits on improvement of performance. In Proceedings of the International Symposium on Performance Science, Utrecht, The Netherlands,15–18 December 2009 pp. 11–21.