Article

Gamifying Physics Laboratory Work Increases Motivation and Enhances Acquisition of the Skills Required for Application of the Scientific Method

Ana Okariz [1,*], Marta Huebra [1], Ane Sarasola [2], Julen Ibarretxe [1], Gorka Bidegain [3,4] and Jose Luis Zubimendi [1]

[1] Department of Applied Physics, Faculty of Engineering in Bilbao, University of the Basque Country (UPV/EHU), Paseo Rafael Moreno "Pitxitxi" 3, 48013 Bibao, Spain; marta.huebra@ehu.eus (M.H.); julen.ibarretxe@ehu.eus (J.I.)

[2] Department of Applied Physics, Faculty of Engineering, Gipuzkoa, University of the Basque Country (UPV/EHU), Europa Plaza 1, 20018 Donostia, Spain; ane.sarasola@ehu.eus

[3] Department of Applied Mathematics, Faculty of Engineering in Bilbao, University of the Basque Country (UPV/EHU), Paseo Rafael Moreno "Pitxitxi" 3, 48013 Bibao, Spain; gorka.bidegain@ehu.eus

[4] Research Centre for Experimental Marine Biology & Biotechnology, Plentzia Marine Station, University of the Basque Country (PiE-UPV/EHU), Areatza Pasealekua, 48620 Plentzia, Spain

[*] Correspondence: ana.okariz@ehu.eus

Abstract: Although increasing student motivation is widely accepted to enhance learning outcomes, this relationship has scarcely been studied quantitatively. Therefore, this study aimed to address this knowledge gap by exploring the effects of gamification on students' motivation and consequently their learning performance, regarding the proper application of the scientific method. To motivate students and enhance their acquisition of new skills, we developed a gamification framework for the laboratory sessions of first-year physics in an engineering degree. Data regarding student motivation were collected through a Likert-scale-type satisfaction questionnaire. The inter-item correlations and Cronbach's alpha confirmed the internal consistency of the questionnaire. In addition, the learning outcome was assessed based on the students' laboratory reports. Students participating in gamified activities were more motivated than those participating in non-gamified activities and obtained better learning results overall. Our findings suggest that gamified laboratory sessions boost students´ extrinsic motivation, and consequently inspire their intrinsic motivation and increase their learning performance. Finally, we discuss our results, with a focus on specific skills and the short- and long-term effects of gamification.

Keywords: gamification; motivation; learning outcome; scientific skills

1. Introduction

A common definition of gamification is "the use of game design elements in non-game contexts" [1]. In general, the term is used if game elements and mechanics are introduced to motivate the end-users of an activity [2]. The ultimate goal is to stimulate participants' action and promote their use of the skills required for the game in different contexts. In educational environments, in particular, adoption of gamified experiences has quickly spread, because of the potential to engage students and capture their attention [3,4]. Furthermore, gamification can meet the demand for new methods that adapt learning processes to "digital native" students [5]. Many studies in the literature have reported the multiple benefits of gamification, such as helping students internalize a topic on a deeper level; providing immediate feedback, incorporating a role-playing aspect that brings students into new environments and situations; and empowering students when they achieve the goals of the given activities [6–8].

A strong argument supporting gamification of the teaching process is based on the hypothesis that gamification significantly increases motivation, thus improving learning

results. A recent study has demonstrated that motivation is highly dependent on the design of innovative and updated topics that engage students and foster better learning attitudes, as well as on the use of new methods that stimulate knowledge acquisition [9]. The author of [10], in a seminal study, has unified the results reported in the literature regarding serious games and gamification. Serious games are defined in the educational context as "games in which education (in its various forms) is the primary goal, rather than entertainment" [11]. Studies about gamification and learning have proposed that behaviors and attitudes associated with the instructional content may affect the learning outcomes. The authors of [12] explained this same idea by classifying the outcomes into experiential (psychological ones) and instrumental (learning ones) and [13] reviewed the literature, trying to create a quantitative scale to evaluate the learning outcomes. The aim of gamification is to change the behaviors and attitudes relevant to learning, and consequently improve learning results. Self-determination theory has been widely applied to gamification [14,15]. The findings have indicated that satisfying the needs for competence, autonomy and social relatedness is also central to intrinsic motivation, and that enriching the learning environments with such elements can affect learning outcome through modifying said environments [16].

However, recent reviews have noted that additional research is still necessary to thoroughly assess the beneficial outcomes claimed in most prior studies. The authors of [17], in a critical review, have reported that many studies on this topic have not presented conclusive evidence of the results and, disagreement exists among the studies following rigorous procedures: although most such studies have indicated beneficial effects of gamification, several have demonstrated negative effects. A meta-analysis by [18] has indicated that the factors contributing to successful gamification remain unknown. Despite considerable research efforts in this field [2,19], a conclusive meta-analysis indicating the effectiveness of gamification in the context of learning and education has yet to be reported. Further high-quality research is needed to enable a more conclusive investigation of the relationship between gamification and learning. Although a diversity of gamification dynamics already exists [20–23], tending towards an individually tailored gamification scheme [24], gamified learning techniques continue to evolve, and new instructional mechanisms are still developed and designed. However, reviews focusing on different aspects related to gamification agree on the need for a unified research-based theory of gamified learning [13,25].

The main objective of this work was to partially fill this knowledge gap through rigorous quantitative research on the correlation between motivation and learning. A comparative study of the learning results of students trained with and without gamification activities was performed. In addition, the motivation of the students was evaluated. Different gamification activities based on a competition–collaboration scheme were designed, because the social dynamics appears to favor more efficient, mainly procedural learning [18]. Importantly, our analysis was based on a combination of cognitive and procedural learning, which is not very common in the literature.

2. Materials and Methods

The correlation between students' satisfaction and learning outcomes was studied, and both aspects were quantified. In the following subsections, after presenting the context of this work, the design of the gamification of the laboratory sessions is described in detail. Subsequently, the procedure for assessing learning outcomes, as well as the survey and procedure designed for assessing student satisfaction, are explained.

2.1. Participants and Context

The present study was performed at the University of the Basque Country, in Bilbao (Spain). All participants were first-year engineering students attending the laboratory sessions of the Introductory Physics for Engineering course (see *Syllabus* in Appendix A). In these sessions, the students worked in groups of three to five (hereafter denoted laboratory groups). Eight laboratory sessions were conducted per academic year (four two-hour sessions/semester, a session per month), with the first three acting as introductory sessions.

In the subsequent sessions (non-introductory sessions), the students were asked to study a physics-associated phenomenon based on defined and limited previous knowledge: they were required to apply the scientific method, starting from the step of hypothesis formulation, and to complete the entire procedure by submitting a report of their work. Each laboratory group submitted a report per laboratory session and received the corresponding feedback and a score. A major goal of the laboratory sessions is for students to acquire the skills required to correctly apply the scientific method, including the necessary data analysis concepts, and to produce proper reports of their investigation. Many studies have reported the benefits of experimental or open-ended methods in promoting the acquisition of such skills [26–30]. During the 2019/20 academic year, two types of learning dynamics were applied to implement these open-ended methods in two randomly formed groups: gamified learning (G group with 136 students) and non-gamified learning (NG group, the control group with 120 students).

2.2. Description of Gamified and Non-Gamified Dynamics

Novice users often have trouble in comprehending the scientific method [31]. In the past few years, deficiencies in the application of the scientific method and data analysis have been reported. Therefore, our instructor team initially implemented the non-gamified laboratory teaching method described in [30,32], consisting of two introductory sessions addressing these deficiencies.

In the first introductory session, the scientific method was explained, and the focus was on the formulation of hypotheses and experimental design. Students were required to complete several exercises (selecting valid hypotheses from a list, proposing an experimental design for a given hypothesis, and finally proposing a hypothesis and its corresponding experimental design), based on the reading material that they were given at the beginning of the course.

The second introductory session focused on the analysis of experimental data, including aspects regarding experimental errors, selection of meaningful methods for plotting data, analysis of the graphical representations and performing adequate regressions. Again, the students were required to complete several data-analysis related tasks within the scope of a given investigation, including the procedures that would be required in the subsequent laboratory sessions. This non-gamified instruction in laboratory methods was used in the control laboratory group in this study (group NG).

The gamification discussed here was applied in two stages. First, the introductory sessions were gamified through the introduction of a leaderboard as a game element in each session. Second, a ranking of the laboratory groups during the entire course was performed to encourage competition, wherein collaboration among teammates is essential to achieve a shared objective [33]. Group G was trained using these methods.

Three introductory sessions were carried out for the G group. The first session covered the graphical representation of data and qualitative and quantitative analysis of graphs. The students were given a fictional context, in which they were required to assume the roles of technicians at an amusement park and analyze whether a Ferris wheel functioned according to its technical specifications. The laboratory work was based on a partially incomplete experimental data set (i.e., a table consisting of position-time values in which some parts of the headers and the data were missing, and therefore it was not obvious what kind of data it corresponded to). The students were required to deduce what physical magnitude was hidden in the data in the tables and determine the specifics of the physical phenomenon that was being investigated through the quantitative analysis of the graphical representations of the data. The students were guided through the exercise. To ensure that students were not stuck at any particular step, hints were provided for the most difficult tasks; these hints were programmed to be displayed on a screen after the estimated time required to perform the corresponding task had elapsed. The role-playing aspect of this activity and the deductive type of guesswork that students were required to perform to interpret the given incomplete data sets were part of the gamification designed for this

laboratory session. Faster completion of a task was awarded higher scores if the results were correct.

The second gamified introductory session covered the handling of experimental errors. The students were given a new fictional context, in which they were asked to assume the role of scientists. Here, the students completed a series of calculations based on the experimentally measured data for a given investigation. Again, programmed tips were used to ensure that all groups completed the assignment in time. In this case, the students received immediate feedback after each task, and were allowed to reattempt the tasks until they were correctly completed. More correct responses and faster responses received higher scores.

Finally, a third introductory session covered the formulation of hypotheses and experimental design. Students were provided with an online quiz, in which they were required to select the correct hypothesis or experimental design among several proposals in different situations. Selecting the correct response for one situation would unblock the next one. Faster completion of tasks resulted in higher scores.

After all three introductory sessions were completed, each group received a single score, obtained as the sum of the three introductory sessions' scores. This single score determined the initial position in the ranking. The ranking position of each laboratory group varied throughout the course, according to the score obtained in the reports of the subsequent laboratory sessions (non-introductory sessions).

2.3. Assessment of Student Satisfaction through a Survey

All participants were asked to rate their satisfaction with different aspects of the learning activity by using a suitable measuring instrument based on the work of [34,35]. The instrument consisted of a survey composed of 16 items (questions) in three categories: (M) motivation, (C) commitment and attitude and (A) subject-associated knowledge, materials and suitability (see Table A1 in Appendix B). The items were evaluated on a five-point Likert scale [36]: (1) strongly disagree, (2) disagree, (3) neutral, (4) agree and (5) strongly agree. These Likert-type items were combined into a single composite score/variable for each of the main themes, rating student satisfaction. Combined items provide a measure of a personality trait [37].

2.4. Assessment of Learning Outcomes

We evaluated the effects of the gamification of the introductory sessions and of maintaining a ranking for the entire course. A rubric consisting of 40 items was used to assess the reports submitted by the students after each non-introductory session. A total of 21 of the 40 items were related to the skills covered in the introductory sessions in both G and NG groups, and those 21 items were selected to evaluate the learning outcome. Table 1 summarizes the skills evaluated and the number of items for each category (see the list of all items in Table A2, Appendix B). For each item, the only two options were "OK" (indicating that the students showed a satisfactory degree of achievement for that particular skill) or "not OK" (when a particular skill was not sufficiently developed). The success percentage in a group was measured as the percentage of the total number of OK scores for each item.

2.5. Data Analysis
2.5.1. Analysis of the Student Satisfaction Survey
Validity and Internal Consistency

The validity and internal consistency of the Likert scale questionnaire refers to the general agreement between multiple Likert-scale items that composed a composite score of the survey measurement for each construct. Validity was determined by application of a correlation matrix for the variables or constructs used to combine items (M, C and A), and Cronbach's alpha [38] was used to assess internal consistency.

Table 1. Summary of the skill categories used for the evaluation of the effect of the gamification. The collected data correspond to the reports submitted after the first (23 of which were submitted by group G and 20 by group NG) and the second (for which 10 reports came from group G and 15 from group NG) non-introductory sessions. It is common for a significant fraction of the students to drop out during the course, which explains the lower number of data collected after the second non-introductory laboratory session. A detailed list of the 21 items used for the evaluation of the effect of the gamification is shown in Table A2, Appendix B.

Item Category	Evaluated Skill	Number of Items
Hypotheses	Proposal of hypotheses	3
	Formulation of hypotheses	2
Experimental design and setup	Experimental design	4
	Experimental setup	2
Analysis of the results	Presentation of experimental conditions and collected data	2
	Graphical representation and quantification	5
	Error handling	1
	Adequate expression of the measurements	2

The normality Shapiro–Wilk test indicated that the data distribution significantly differed from normality. Consequently, the non-parametric rank-based Spearman correlation test was applied for the correlation matrix. This matrix showed the inter-item score correlation and the item total correlation, thus yielding a total score for each item and then correlating each item's score with the total score.

The traditional alpha coefficient [38] may not be robust to the violation of the normal assumption and missing data. Non-normal data tend to result in additional error or bias in estimating internal-consistency reliability, thus making the alpha coefficient of the sample less accurate [39]. Consequently, the robust Cronbach alpha [40] was used to explore the reliability or internal consistency of the questionnaire. With this method, robust confidence intervals were estimated for the alpha coefficient. In addition, the alpha coefficient was estimated through removal of items for each of the three variables measuring satisfaction.

Effects of Gamification Sessions on Student Satisfaction

The survey was conducted among students trained in the traditional and gamified sessions. The effect of the gamified sessions on students' satisfaction with the learning experience was tested with the Wilcoxon rank-sum test with continuity correction, because the data distribution significantly differed from normality. Similarly, the effects of gender and course repetition on the students' satisfaction with the learning experience were explored with the same test.

2.5.2. Analysis of Learning Outcomes

The data from the students' evaluations of both gamified and non-gamified sessions were not normally distributed. The 2×2 contingency and expected frequency tables were created for the categories presented in Table 1. Because the sample was small, and/or the expected counts were lower than 5% in some cases, the independence between the factors of gamification and learning outcome (success) was tested with nonparametric Fisher's exact test [41] instead of the more common chi-square test.

The data were analyzed in R 4.0.3 software [42]. Specifically, the internal consistency of the questionnaire was explored with the R package alpha coefficient. The following R packages were also used to generate the results presented: psych, corrplot, corr, multiplot, dplyr, PerformanceAnalytics and Hmisc.

3. Results

3.1. Satisfaction, Survey Validity and Internal Consistency

Overall, inter-item and item total correlations were significant for the categories M, A and C, which showed moderate and strong positive correlations, particularly with the total average score (see Figure 1 for category M, and Figures A1 and A2 in Appendix B for categories C and A, respectively). The non-linear fitted curves overall indicated a repetitive pattern of positive relationships. The items in category M showed the highest significant correlation among items and with the total average score. Items M3 and M4 showed the highest correlation. For categories A and C, the inter-item and item total correlations were also significant, and moderate positive correlations were observed for most items.
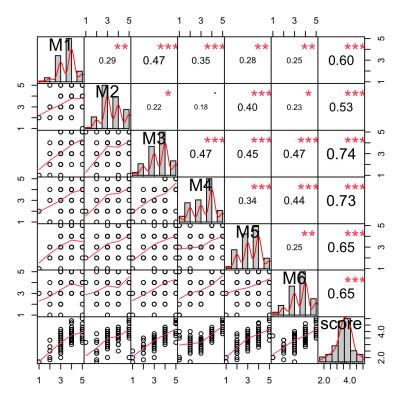

Figure 1. The frequency distribution of scores for each item response in category M, both individually (M1–M6) and combined (score), is presented on the diagonal. Below the diagonal, the bivariate scatter plots with a fitted non-linear curve are displayed. Above the diagonal, the value of the inter-item correlation and the correlation of each item with the total score (Spearman rank) are plotted. The significance level is indicated with symbols "***", "**", "*" and ".", denoting *p*-values of 0.001, 0.01, 0.05 and 0.1, respectively.

The robust Cronbach alpha with 95% confidence intervals was estimated for all items, together and separately, for each of the categories. The alpha coefficient was 0.89 (0.84, 0.94) (average alpha (Lower CI, Upper CI)) for all items, thus suggesting a high overall internal consistency of the questionnaire. The internal consistency of category M was the highest among all three categories, with an alpha coefficient of 0.87 (0.81, 0.93).

The alpha coefficient was also estimated for each category after removing, in turn, each item of the category (see Table 2 for category M and Table A3 in Appendix B for categories C and A). For category M, removing items resulted in a maximum increase in the average (whole M-scale) alpha, from an alpha of 0.87 to 0.88. This increase was not statistically significant, given the confidence interval and standard error; that is, for this category, none of the items were redundant or unnecessary.

Table 2. Average coefficient alpha, CI and SE if an item is dropped for category M. Rows M1–M6 present coefficient alpha when that item is dropped from the questionnaire. The last row, M, presents the value of alpha for the whole scale of the category, with no dropped items.

	Items M			
	alpha	**Lower CI**	**Upper CI**	**SE**
M1	0.85	0.75	0.95	0.05
M2	0.88	0.79	0.97	0.04
M3	0.83	0.70	0.96	0.06
M4	0.86	0.76	0.96	0.05
M5	0.85	0.74	0.96	0.05
M6	0.87	0.78	0.96	0.05
M	0.87	0.81	0.93	0.03

3.2. Effects of Gamified Sessions on Student Motivation

Student motivation was assessed using the answers to category M of the satisfaction questionnaire. The average scores obtained in this category were 3.71 ± 0.76 (mean $\pm$ SD) for group G and 3.40 ± 0.55 for group NG. The effect of gamified sessions on motivation was significant (Wilcoxon rank-sum test: $W = 1617$, $p = 0.002$).

The effects of gender and course repetition on student satisfaction were tested to detect additional factors affecting student satisfaction. No effect of these two factors was observed ($W = 88$, p-value $= 0.8741$ for gender and $W = 100$, p-value $= 0.831$ for course repetition).

3.3. Effects of Gamified Sessions on Learning Outcomes

Several significant differences were observed for groups G and NG, based on the results shown in Table 3. According to the data corresponding to the first session report, the students from group G obtained more satisfactory results in the formulation of hypotheses, specifically for the items assessing the validity of the hypotheses and their correct formulation (Table A2 in Appendix B). Moreover, they also obtained better results in the "analysis of results" section, and this effect was particularly significant for the calculation of errors and expression of measurements.

The differences between groups G and NG were less relevant in the second-session reports and the categories that showed significant differences in the first report became less differentiated. However, new significant differences were detected in the categories of experimental design and graphical representation and quantification.

Table 3. Fisher's exact test results (p-value) and the percentage of students with successful results. Results for groups G (trained with a gamified methodology) and NG (trained with a traditional non-gamified methodology) in both first and second laboratory reports. A p-value < 0.05 (in bold *) indicates a significant association between the teaching method and learning success.

	First Laboratory Report			Second Laboratory Report		
	Group G	**Group NG**	**p-Value**	**Group G**	**Group NG**	**p-Value**
Hypotheses						
Proposal of hypotheses (3 items)	100%	91%	0.044 *	96%	100%	0.50
Formulation of hypotheses (2 items)	98%	89%	0.046 *	97%	82%	0.08

Table 3. *Cont.*

	First Laboratory Report			Second Laboratory Report		
	Group G	**Group NG**	***p*-Value**	**Group G**	**Group NG**	***p*-Value**
Experimental design and setup						
Experimental design (4 items)	48%	43%	0.88	69%	46%	0.03 *
Experimental setup (2 items)	34%	15%	1.0	45%	42%	1.0
Analysis of results						
Presentation of experimental conditions and collected data (2 items)	72%	48%	0.27	85%	70%	0.23
Graphical representation and quantification (5 items)	69%	62%	0.88	64%	56%	0.04 *
Error handling (1 item)	30%	5%	0.05 *	29%	27%	1.0
Adequate expression of measurements (2 items)	47%	22%	0.007 *	58%	33%	0.08

4. Discussion and Conclusions

The purpose of this study was to analyze whether a correlation exists between greater motivation among students in a first-year university physics course in an engineering degree program and the learning outcomes concerning the application of the scientific method in laboratory practical examinations. A gamification framework was designed and applied, to achieve higher student motivation. The core of the gamified activities was applied in the initial introductory sessions, and the motivation and learning outcomes were assessed thereafter. The G group received one more introductory session (three sessions) than the NG group (two sessions) due to the characteristics of the game-based dynamic, which required some more time for introducing the same concepts. Since the same concepts were introduced, we do not think this introduces a significant bias into the study. However, we should use caution when interpreting the results of learning performance, and recognize that an extra session for the G group could in part be benefiting the learning process and outcomes

Student satisfaction was evaluated with a Likert-scale-type survey. Inter-item correlations and robust Cronbach's alpha analysis, including deletion diagnostic testing [39], confirmed the validity and reliability of the questionnaire, particularly for the motivation category. The analysis indicated that motivation was higher among students in the G group. Indeed, gamification has been found to be related to extrinsic motivation, and to inspire intrinsic motivation [2,43]. Ref. [2] explored the relationship between gamification and motivation by reviewing 32 studies, 20 of which suggested that gamification increases student motivation, as observed herein. Specifically, in higher education, our results coincide with those of reports suggesting that gamification has positive effects on intrinsic motivation [18,44].

Regarding the learning outcomes, our results suggest that the higher motivation achieved through gamification activities has an overall positive effect on the acquisition of the analyzed skills by students. In general, the success rates (% of students with successful results) achieved by the G and NG groups (shown in Table 3) evolved differently. The

gamification produced an immediate positive impact on the success rates (first report), while this effect was less evident in the long term (second report). The success rates show a slower and more progressive improvement in the NG group. However, the results show an overall sizable effect of gamified activities on longer-term learning performance. This is in line with the conclusions derived from a meta-analysis of 24 studies conducted in Ref. [45]. The less-evident effect of gamification observed in the second report and the potentially long-term decrease in learning performance needs to be carefully investigated. Specifically, the leaderboard constructed to motivate students in the gamified learning environment is a critical step, since these boards can have some limitations associated with their design. In gamified learning environments such as the one here, students receive feedback on their exercises through a leaderboard, and they are rewarded for their accomplishments [46]. Sometimes the gap between learners in terms of learning performance can be wide, and learning motivation decreases, particularly for those in the lower ranks when their achievements are compared with those of the high-ranked learners [47]. This potentially negative effect of the leaderboards can be successfully addressed by applying some design principles that minimize relative deprivation and learners' experiences of failure and maximize learners' experience of success [46]

A parallel reading of the results shown in Table 3 can be achieved through analyzing the statistical significance (*p*-values < 0.05, marked with an asterisk) of the items associated with the skills evaluated in the reports. The significant values demonstrated a positive association between gamification and learning success in the corresponding skill. The results of the first report showed significant differences between the G and NG groups in the categories of hypotheses, error handling and adequate expression of measurements. In the second report, significant differences were found in two additional categories: experimental design, and graphical representation and quantification. Only the skills associated with the experimental setup, and the presentation of the experimental conditions and collected data, did not show any significant differences in learning success.

Despite the significant improvement in learning outcomes, due to an improved motivation through gamification, the success rates for some evaluated skills were not as high as would be desirable. A redesign of the gamified activities directed toward those skills may help solve this problem. The authors of [45] proposed studying the implementation of long-term gamification to explore potential novel effects and for the consolidation and perpetuation of learning outcomes. Further work should be performed to study this issue in depth.

Appendix A

The syllabus of the Introductory Physics course is as follows:
- Dynamics of the particle
- Work, energy and its conservation
- Lineal momentum and its conservation
- Dynamics of the rigid solid
- Angular momentum and its conservation
- Electric field and Gauss's law
- Electric potential
- Capacitors and capacitance
- Direct-current circuits
- Magnetic field
- Electromagnetic induction
- Equations of Maxwell, electromagnetic waves.

Appendix B

Table A1. The Likert Scale questionnaire used in the student satisfaction survey stands for (1) strongly disagree, (2) in disagreement, (3) neutral, (4) in agreement and (5) strongly agree. Items have been arranged in the table according to the three categories analyzed: (M) motivation, (C) commitment and attitude, and (A) subject-related knowledge, materials and suitability.

Motivation (M)				Likert Scale	
I´ve acquired skills invaluable for my training as an Engineer.	1	2	3	4	5
Completing the tasks satisfies me	1	2	3	4	5
The laboratory sessions have been stimulating and enriching	1	2	3	4	5
The laboratory dynamics don't stimulate motivation to learn	1	2	3	4	5
I have progressed during the sessions, which encouraged me to persevere in my efforts	1	2	3	4	5
I get bored during lab sessions	1	2	3	4	5
Commitment and attitude (C)				**Likert scale**	
I feel identified with the way the class is run	1	2	3	4	5
I get involved and try to learn from and teach others	1	2	3	4	5
I am not consistent when faced with the difficulties of the proposed tasks	1	2	3	4	5
My behavior is positive with respect to the development of sessions and tasks	1	2	3	4	5
I would prefer to do the practicals/practical exercises using a different methodology	1	2	3	4	5
I can rely on my teammates for support when I need help	1	2	3	4	5
Subject-related knowledge, materials and suitability (A)				**Likert scale**	
My preparation for this type of lab-work has been adequate	1	2	3	4	5
The methodology applied is appropriate to the characteristics of the groups and to the subject	1	2	3	4	5
I find help and support in the materials provided and/or from the lab instructors	1	2	3	4	5
There is hardly any relationship between the lectures and the laboratory sessions	1	2	3	4	5

Table A2. This Table lists the skills that were trained during the introductory sessions (middle column) and the corresponding possible failed items (right column) that were analyzed to assess the degree of development of those skills. For each item, an OK (indicating that the students showed a satisfactory degree of achievement for that particular skill) or NOT OK (when a particular skill was not sufficiently developed) assessment would be issued, depending on the degree of achievement shown for that particular item. The percentage of success in a group is measured as the percentage of total number of OK marks in each item.

Category of the Items	Evaluated Skill	Item
Hypotheses	Hypothesis proposals	• The proposal is not valid as a hypothesis • The proposed hypothesis is not verifiable with the available material • The proposed hypothesis is not in line with the purpose of the work
	Formulation of hypotheses	• The proposal consists of several hypotheses • The hypothesis is followed by or interspersed with comments
Experimental design and setup	Experimental design	• The magnitudes to be measured are not correctly indicated • Fixed and variable magnitudes are not specified • No analysis of the convenience of repeating measurements • The obtaining of indirect measures is not specified
	Experimental setup	• No explanations and/or diagrams • No details of equipment used and/or its sensitivity
Analysis of results	Presentation of experimental conditions and collected data	• The experimental conditions are not detailed • Measures and measurements are not clearly shown
	Graphical representation and quantification	• Graphs: the axes lack adequate information • Graphs: no qualitative analysis of the relationship between magnitudes • Graphs: no data fitting or not used • Incorrect calculation of measurements obtained from the fitting • Incorrect calculation of measurements not obtained from the fitting
	Error handling	• Most errors are calculated correctly
	Adequate expression of measurements	• Most errors are not expressed correctly • No units have been indicated

Table A3. Average coefficient alpha, CI and SE if an item is dropped for categories A and C. For both categories, dropping items did not result in a statistically significant increase in the whole-scale alpha with respect to the current alpha of 0.73 (0.6, 0.88) for C and 0.72 (0.58, 0.87) for A, respectively. The last row represents the values for the whole scale of the category with no dropped items. Similarly, with category M, there is no need to drop any item from the questionnaire for these two categories.

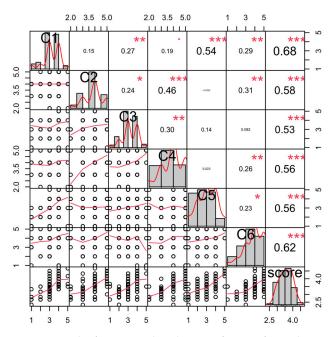|	Category C					Category A			
	Alpha	Lower CI	Upper CI	SE		Alpha	Lower CI	Upper CI	SE
C1	0.70	0.48	0.92	0.11	**A1**	0.77	0.54	1.00	0.12
C2	0.73	0.54	0.92	0.10	**A2**	0.62	0.23	1.00	0.20
C3	0.70	0.48	0.92	0.11	**A3**	0.65	0.29	1.00	0.18
C4	0.69	0.46	0.92	0.12	**A4**	0.77	0.54	1.00	0.11
C5	0.74	0.55	0.93	0.10					
C6	0.72	0.52	0.92	0.10					
C	0.73	0.60	0.88	0.07	**A**	0.72	0.58	0.87	0.07

Figure A1. The frequency distribution of scores for each item response in category C and the overall score is presented on the diagonal. Below the diagonal, bivariate scatter plots with a fitted non-linear curve are displayed. The value of the inter-item correlation and the correlation of each item with the total score (Spearman Rank), and significance level represented by stars, are plotted above the diagonal. Each significance level is associated with a symbol: *p*-values 0.001, 0.01, 0.05, 0.1 and 1 are associated with symbols "***", "**", "*", "." and " ", respectively.
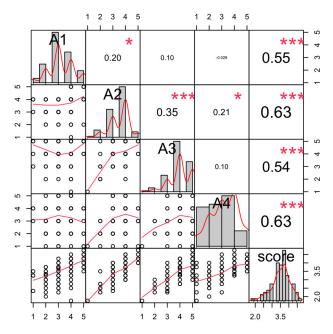

Figure A2. The frequency distribution of scores for each item response in category A and the overall score is presented on the diagonal. Below the diagonal, bivariate scatter plots with a fitted non-linear curve are displayed. The value of the inter-item correlation and the correlation of each item with the total score (Spearman Rank), and significance level represented by stars, are plotted above the diagonal. Each significance level is associated with a symbol: *p*-values 0.001, 0.05, and 1 are associated with symbols "***", "*", and " ", respectively.

References

1. Deterding, S.; Dixon, D.; Khaled, R.; Nacke, L. From game design elements to gamefulness: Defining "gamification. In Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, Tampere, Finland, 28–30 September 2011; New York Association for Computing Machinery Publishing: New York, NY, USA, 2011; pp. 9–15. [CrossRef]
2. Seaborn, K.; Fels, D.I. Gamification in theory and action: A survey. *Int. J. Hum.-Comput. Stud.* **2015**, *74*, 14–31. [CrossRef]
3. Linehan, C.; Kirman, B.; Lawson, S.; Chan, G. Practical, appropriate, empirically-validated guidelines for designing educational games. In Proceedings of the ACM Annual Conference on Human Factors in Computing Systems, Vancouver, BC, Canada, 7–12 May 2011; New York Association for Computing Machinery Publishing: New York, NY, USA, 2011; pp. 1979–1988.
4. Bonde, M.T.; Makransky, G.; Wandall, J.; Larsen, M.V.; Morsing, M.; Jarmer, H.; Sommer, M.O.A. Improving biotech education through gamified laboratory simulations. *Nat. Biotechnol.* **2014**, *32*, 694–697. [CrossRef]
5. Prensky, M. Digital Natives, Digital Immigrants. *Horiz* **2001**, *9*, 1–6.
6. Gee, J.P. Video Games and Learning. Week 1. Video 6/8. 13 Principles of Game-Based Learning. Available online: www.youtube.com/watch?v=bLdbIT-exMU&list=PLl02cFD2W03xWmCDnf_N78bIpMkfYEck (accessed on 28 October 2020).
7. Lo, C.K.; Hew, K.F. 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]
8. Larson, K. Serious Games and Gamification in the Corporate Training Environment: A Literature Review. *Techtrends* **2020**, *64*, 319–328. [CrossRef]
9. Fuhrmann, T. Motivation Centered Learning. In Proceedings of the IEEE Frontiers in Education Conference (FIE), San José, CA, USA, 3–6 October 2018.
10. Landers, R.N. Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simul. Gaming* **2014**, *45*, 752–768. [CrossRef]
11. Michael, D.; Chen, S. *Serious Games: Games That Educate, Train, and Inform*; Thomson Course Technology: Boston, MA, USA, 2005.
12. Liu, D.; Santhanam, R.; Webster, J. Toward Meaningful Engagement: A Framework for Design and Research of Gamified Information Systems. *MIS Q.* **2017**, *41*, 1011–1034. [CrossRef]
13. Inocencio, F. Using gamification in education: A systematic literature review. In Proceedings of the Thirty Ninth International Conference on Information Systems, San Francisco, CA, USA, 13–16 December 2018.
14. Mekler, E.D.; Brühlmann, F.; Tuch, A.N.; Opwis, K. Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Comput. Hum. Behav.* **2017**, *71*, 525–534. [CrossRef]
15. 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]
16. Ryan, R.M.; Deci, E.L. Overview of self-determination theory: An organismic dialectical perspective. In *Handbook of Self-Determination Research*; Ryan, R.M., Deci, E.L., Eds.; University of Rochester Press: Rochester, NY, USA, 2002; pp. 3–33.
17. Dichev, C.; Dicheva, D. Gamifying education: What is known, what is believed and what remains uncertain: A critical review. *Int. J. Educ. Technol. High. Educ.* **2017**, *14*, 9. [CrossRef]
18. Sailer, M.; Homner, L. The Gamification of Learning: A Meta-analysis. *Educ. Psychol. Rev.* **2020**, *32*, 77–112. [CrossRef]
19. Hamari, J.; Koivisto, J.; Sarsa, H. Does gamification work?—A literature review of empirical studies on gamification. In Proceedings of the 47th Annual Hawaii International Conference on System Sciences (ed Sprague RH Jr), Waikoloa, HI, USA, 6–9 January 2014; IEEE: Washington, DC, USA, 2014; pp. 3025–3034. [CrossRef]
20. Kapp, K.M. *The Gamification of Learning and Instruction*; Pfeiffer: San Francisco, CA, USA, 2012.
21. Kyewski, E.; Krämer, N.C. To gamify or not to gamify? An experimental field study of the influence of badges on motivation, activity, and performance in an online learning course. *Comput. Educ.* **2018**, *118*, 25–37. [CrossRef]
22. Rapp, A.; Hopfgartner, F.; Hamari, J.; Linehan, C.; Cena, F. Strengthening gamification studies: Current trends and future opportunities of gamification research. *Int. J. Hum.-Comput. Stud.* **2019**, *127*, 1–6. [CrossRef]
23. Tsay, C.H.-H.; Kofinas, A.; Luo, J. Enhancing student learning experience with technology-mediated gamification: An empirical study. *Comput. Educ.* **2018**, *121*, 1–17. [CrossRef]
24. Oliveira, W.; Hamari, J.; Shi, L.; Toda, A.M.; Rodrigues, L.; Palomino, P.T.; Isotani, S. Tailored gamification in education: A literature review and future agenda. *Educ. Inf. Technol.* **2022**, *28*, 373–406. [CrossRef]
25. Zainuddin, Z.; Chu, S.K.W.; Shujahat, M.; Perera, C.J. The impact of gamification on learning and instruction: A systematic review of empirical evidence. *Educ. Res. Rev.* **2020**, *30*, 100326. [CrossRef]
26. Gott, R.; Duggan, S.; Johnson, P. What do Practising Applied Scientists do and What are the Implications for Science Education? *Res. Sci. Technol. Educ.* **1999**, *17*, 97–107. [CrossRef]
27. Lottero-Perdue, P.S.; Brickhouse, N.W. Learning on the job: The acquisition of scientific competence. *Sci. Educ.* **2002**, *86*, 756–782. [CrossRef]
28. Bybee, R.W.; Fuchs, B. Preparing the 21st century workforce: A new reform in science and technology education. *J. Res. Sci. Teach.* **2006**, *43*, 349–352. [CrossRef]
29. Hofstein, A.; Kind, P.M. Learning in and from science laboratories. In *Second International Handbook of Science Education*; Fraser, B.J., Tobin, K.G., McRobbie, C.J., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 189–207.

30. Sarasola, A.; Rojas, J.F.; Okariz, A. Training to Use the Scientific Method in a First-Year Physics Laboratory: A Case Study. *J. Sci. Educ. Technol.* **2015**, *24*, 595–609. [CrossRef]

31. Kirschner, P.; Meester, M. Laboratory approaches. *High. Educ.* **1988**, *17*, 81–98. [CrossRef]

32. Huebra, M.; Ibarretxe, J.; Okariz, A.; Sarasola, A.; Zubimendi, J.L. Game-based learning of scientific skills. In Proceedings of the Edulearn 20 Proceedings, Online, 6–7 July 2020; pp. 2052–2057. [CrossRef]

33. Vegt, N.; Visch, V.; Ridder, H.; Vermeeren, A. Designing Gamification to Guide Competitive and Cooperative Behaviour in Teamwork. In *Gamification in Education and Business*; Reiners, T., Wood, L.C., Eds.; Springer: Cham, Switzerland, 2015; pp. 513–533.

34. Weiner, B. Intrapersonal and Interpersonal Theories of Motivation from an Attributional Perspective. *Educ. Psychol. Rev.* **2000**, *12*, 1–14. [CrossRef]

35. Biggs, J.; Kember, D.; Leung, D.Y. The revised two-factor Study Process Questionnaire: R-SPQ-2F. *Br. J. Educ. Psychol.* **2001**, *71*, 133–149. [CrossRef] [PubMed]

36. Likert, R. A technique for measurement of attitudes. *Arch. Psychol.* **1932**, *140*, 5–55.

37. Boone, H.N.; Boone, D.A. Analyzing Likert Data. *J. Ext.* **2012**, *50*, 1–5.

38. Cronbach, L.J. Coefficient alpha and the internal structure of tests. *Psychometrika* **1951**, *16*, 297–334. [CrossRef]

39. Sheng, Y.; Sheng, Z. Is Coefficient Alpha Robust to Non-Normal Data? *Front. Psychol.* **2012**, *3*, 34. [CrossRef] [PubMed]

40. Zhang, Z.; Yuan, K.-H. Robust Coefficients Alpha and Omega and Confidence Intervals With Outlying Observations and Missing Data. *Educ. Psychol. Meas.* **2016**, *76*, 387–411. [CrossRef] [PubMed]

41. Fisher, R.A. On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P. *J. R. Stat. Soc.* **1922**, *85*, 87. [CrossRef]

42. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. Available online: http://www.r-project.org/index.html (accessed on 3 April 2020).

43. Hsieh, T.-L. Motivation matters? The relationship among different types of learning motivation, engagement behaviors and learning outcomes of undergraduate students in Taiwan. *High. Educ.* **2014**, *68*, 417–433. [CrossRef]

44. Subhash, S.; Cudney, E.A. Gamified learning in higher education: A systematic review of the literature. *Comput. Hum. Behav.* **2018**, *87*, 192–206. [CrossRef]

45. Bai, S.; Hew, K.F.; Huang, B. Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educ. Res. Rev.* **2020**, *30*, 100322. [CrossRef]

46. Park, S.; Kim, S. Leaderboard Design Principles to Enhance Learning and Motivation in a Gamified Educational Environment: Developmental Study. *JMIR Serious Games* **2021**, *9*, e14746. [CrossRef] [PubMed]

47. Walker, I.; Pettigrew, T. Relative deprivation theory: An overview and conceptual critique. *Br. J. Soc. Psychol.* **1984**, *23*, 301–310. [CrossRef]