*Article*

# Moodle Quizzes as a Continuous Assessment in Higher Education: An Exploratory Approach in Physical Chemistry

**Isabel López-Tocón** [ID]

Departamento de Química Física, Facultad de Ciencias, Campus de Excelencia Andalucía Tech, Universidad de Málaga, E-29071 Málaga, Spain; tocon@uma.es

**Abstract:** The use of Moodle quizzes as a continuous assessment and an integral part of the educational methodology in higher education has been analyzed in a case study of physical chemistry subject. Two types of quiz designed with different item types and different settings, called basic quiz (BQ) and thematic block quiz (TBQ), were elaborated making use of a question bank with more than 450 items. BQ has true/false items, while TBQ has randomly mixed items (multiple choice, numerical and matching). The effect of the type of quiz on the student scores is analyzed according to statistical and psychometric data such as the degree of participation, the facility index and the discrimination index of each item, and the average score, calculated according to the classical test theory. This allows us to discern which type of quiz has an enough quality to use it as an assessment tool. Moreover, the effect of this educational activity, developed during the last six academic years from 2014 to 2020, just before of the pandemic situation, is evaluated considering the scores of the students in the Ordinary Calls of exams and comparing them with previous courses taught with a traditional education based on master classes. The statistic results indicate that TBQs are more discriminative than BQs and could be used as an assessment tool, while BQs could be only useful as formative activity. Moodle quizzes turn out to be a reliable strategy for learning of contents in scientific matter, with a high participation in the knowledge tests, with good marks in the average score and a greater number of pass degrees in the Ordinary Calls.

**Keywords:** moodle quizzes; continuous assessment; statistical analysis; physical chemistry

## 1. Introduction

The adaptation of the subjects of any graduate degree to the new European Space for Higher Education, where new capacities and abilities are evaluated in students, implies a significant change in the traditional teaching methodology that has been developed, mainly as master classes. This change in the pedagogical model of teaching–learning is in turn conditioned and reinforced by the new model of current digital society and by the new information and communication techniques available in any space-time framework [1,2]. The availability of information on everyday electronic devices such as smartphones or tablets, in addition to the connectivity of user groups to the internet, allows establishing other more active work dynamics, giving to the student a greater participation in the teaching–learning process of a subject [3].

The studies carried out to evaluate the global impact of the use of technology on student performance are not conclusive, yielding different results [4,5] given that these teaching researches may depend on other factors not identified in the analysis itself, such as the educational method or strategy [6] that is carried out through the electronic medium, or the student's commitment to the learning methodology [7]. However, these multimedia and interactive technologies can be of great help in offering quality comprehensive education [8] based on current computer tools that facilitate cognitive learning processes and reinforce the capacities of abstract reasoning and study of a specific subject, in addition to complete the traditional forms of learning [9].

Different teaching methodologies integrate the technological devices in the educational environment such as the blended-learning [10–12] and gamification [13–15]. Blended-learning, b-learning, combines face-to-face lessons in the classroom, required for any subject in university study plans, and virtual training teaching activities through learning platforms, while the gamification techniques try to create similar experiences to those experienced when playing games in order to motivate and engage users. Both methodologies profit from the presence of the teacher as a transmitter of knowledge and guide of educational activities and from the communication technology that facilitates independent and collaborative learning. In particular, b-learning has been applied in subjects from different areas of knowledge such as education sciences [12], natural sciences [16], economics [9], engineering [7], etc. as a proposal for European convergence, given that it allows the student's noncontact work hours to be completed with virtual activities, as established in the new university teaching guides for the convergence of the European Space for Higher Education.

Moodle platform [17] is a virtual learning environment that offers very attractive functionalities from the pedagogical point of view by promoting the philosophy of constructivist social education [18,19], and where the subjects can be accommodated with easy handling, at the editing and user level by teachers and students, respectively. In this virtual environment, teaching resources of different characteristics can be included, such as links to web pages, chats, forums, messages, and other specific documents like notes, tutorials and question relationships elaborated by the teacher. Moreover, it offers the possibility of carrying out online activities through quizzes, which could allow the continuous assessment of students' learning. A great variety of quizzes can be designed with different item types and settings, but not all quizzes can differentiate the skills and competences of student, and thus, they could not be used as assessment tools. The quality of these quizzes can be analyzed by statistical and psychometric data reported by Moodle platform [20,21].

Concerning evaluation methods by using online quizzes, there are studies in diverse disciplines such as engineering, biology, medicine and the social sciences [22–24]. Although, there are some objections to the implementation of such systems related to the confidentiality of the identity of the student, the use of the information and its possible impact on the educational process [25,26]; these offer some advantages such as the efficient management of results in a huge students' group, the speed by which the evaluation can be performed, and the save of paper [27]. However, the design of quizzes must be adequately elaborated in order to be used as an assessment tool. Two important points must be considered in the design, such as the writing of different questions using different item type and the own quiz settings. The statistical and psychometric data derived from a particular quiz can a great help us know the quality of the quiz. There are some studies regarding the analysis of information generated from test-type quiz evaluations in other scientific subjects [20,21,28–30], yielding how such results could be useful for professors and students. No statistical and psychometric studies on physical chemistry quizzes have been found in the bibliography.

In this work, two types of Moodle quizzes are designed in physical chemistry subject. The main objective is to establish which type of quiz can be used as an assessment tool on the basis of statistical and psychometric data. Here, it highlights how Moodle statistics can be used to measure the effectiveness and reliability of a quiz. In addition, the effect of these online activities on the final scores of the students are compared with those obtained in a traditional education.

## 2. Materials and Methods

The research is designed in three stages: first, the student population was surveyed by a brief poll to inquire about their entry into the university; second, the students answered the quizzes during the teaching semester; and finally, the statistical and psychometric parameters of the quizzes was analyzed on the basis of the classical test theory [31–33]

(See Supplementary Materials). A brief survey is carried out at the end of the teaching period to know the opinion of the students about this experience. The scores obtained in the two Ordinary Calls of exams are compared with those obtained in previous courses where the teaching methodology corresponds to a traditional education based exclusively on master classes.

This research is performed in the general physical chemistry subject during six years, from the 2014–2015 to the 2019–2020 academic years, just before of the pandemic situation. This matter is included in the Basic Module of the Degree in Chemistry at the University of Málaga. It consists of six theoretical credits, and it is taught during the first semester of the first year of the degree.

This subject was chosen because it is a difficult matter for novel students in the Degree of Chemistry. It includes themes like thermodynamic, electrochemistry and kinetics that are the starting point of other physical chemistry subjects in higher courses, in which a significant dropout of students has been detected. Thus, it seems convenient to apply a new educational methodology, or new activities using technological devices, in the first course in order to consolidate and strengthen the basic concepts of this matter.

### 2.1. Sample

The average number of students in general physical chemistry was about 80 students during the last academic years, with a parity proportion of men and women in the last four years. All students can freely participate in the quizzes as a unique experimental group. No specific sampling method and no control group is established with the aim that all students were evaluated in a homogenous way so that there are no discrepancies in the final evaluation.

It was not possible to perform a similar study in other courses or scientific areas, even in other degrees, because there were no other teachers implied in the project using a similar educational strategy with Moodle quizzes. Although this sample is not representative of the higher education context, the similar results obtained in this experience along different years with different students population point out that it would not be expected to see significant changes in another similar scientific scene, giving probably a similar trend.

At the beginning of the course, a brief survey is carried out to explore the admission at the university, such as academic background on chemistry knowledges and the enrollment in the degree. Considering an average of the last six academic years, practically all the students, 86–88%, are 18 years old, and the rest, 11–12%, are in the range of 21 to 25 years old, which could probably be due to repeaters in the secondary or bachelor cycle, or students who come from other degrees. Most of the students, 86–95%, have studied a chemistry subject during high school, but it should be noted that about 4–5% of students have not studied any chemistry subject in any official degree before to their admission to the university, although they indicate that they have basic knowledge of chemistry. Only a small proportion, 1–2%, have no knowledge of chemistry. Moreover, a high proportion, around 70–85%, has enrolled in the Chemistry degree because it is their vocation, being the first option in university pre-registration. Only 12–25% of students recognize that it is not their vocation and it has not been the first option in the university pre-registration. In addition, this degree was not the first choice of about 1–2% of students, but it was the only option for their admission to the university.

### 2.2. Development of the Experience: Didactic Strategy

Within the Moodle platform, a question bank has been created and divided into five thematic blocks that involve all topics of the teaching program (Table 1). Each block has more than 50 questions or items, even over 100 items in the cases of the Matter and Thermodynamics blocks. The question bank has over 450 items belonging to four types of Moodle questions: true/false, multiple choice (with multi-responses and single response), matching and numerical. All these items were elaborated according to the scientific competencies required for passing this subject.

**Table 1.** Teaching program of the general physical chemistry subject developed in eleven lessons and distributed in five thematic blocks. Available at https://oas.sci.uma.es:8443/ht/2020/ProgramasAsignaturas_Titulacion_5004_AsigUMA_51 635.pdf (accessed on 22 July 2021).

| Theme | Brief Description of the Topic | Thematic Block | | N° Moodle Items |
|---|---|---|---|---|
| 1<br>2<br>3 | The matter: Laws of chemical combination<br>Gaseous state<br>Liquid and solid states | 1. | Matter (102 items) | 27 True/False<br>53 Multiple choice<br>9 Matching<br>10 Numerical |
| 4<br>5 | Solutions<br>Colligative properties of solutions | 2. | Solutions (98 items) | 34 True/False<br>49 Multiple choice<br>6 Matching<br>9 Numerical |
| 6<br>7<br>8 | Zeroth Law and First Law of Thermodynamics<br>Second and Third Law of Thermodynamics<br>Chemical equilibrium | 3. | Thermodynamics (122 items) | 31 True/False<br>62 Multiple choice<br>14 Matching<br>15 Numerical |
| 9<br>10 | Electrolytes and electrolytic solutions<br>Electrochemistry: electrolytic and galvanic cells | 4. | Electrochemistry (84 items) | 20 True/False<br>50 Multiple choice<br>7 Matching<br>7 Numerical |
| 11 | Kinetics chemistry | 5. | Kinetics (53 items) | 13 True/False<br>29 Multiple choice<br>2 Matching<br>9 Numerical |

The set of items is classified, in turn, into two categories, one with the questions that collect the basic knowledge of the subject, while the other contains more elaborate questions, in order to check the skills and abilities of the students in practical reasoning about physical chemistry. In this way, two types of quiz are developed. First, a "basic" quiz (BQ) is proposed for each of the eleven topics. It consists of ten true/false type items, with a time limit of one hour. The BQ contains the same questions for all students and is active for a period of one week after finishing the topic in class. Second, another type of "thematic block" quiz (TBQ) is proposed corresponding to each of the five thematic blocks, which are made up of several topics in the teaching program, except those dedicated to chemistry kinetics (see Table 1). It has ten items of different type (multiple choice, numerical, matching) chosen at random from a category of question bank, so it is practically an individual and different test for each student. The multiple choice items have a particular characteristic, the correct/incorrect answers score positively/negatively, with a proportional value to the number of item options. These quizzes are held in a scheduled day.

In both types of quizzes, each item has the same statistical weight of 10% in the final mark. All quizzes are performed outside the classroom and have a delayed feedback; that is, the correct answers can be only checked once the test is over for all students. All these activities are carried out continuously throughout the semester according to the physical chemistry program.

All students were informed about the characteristic of Moodle quizzes and how the platform works before doing the activities. In this way, any bias factor due to students' attitudes towards technology along the time would be diminished.

### 3. Results and Discussion

*3.1. Participation in the Quizzes*

There is a high participation during the last six academic courses (Figure 1), higher than 50% in any BQ or TBQ quizzes, with the exception of the last BQ carried out in the 2017–2018 academic year with a participation of 40%.
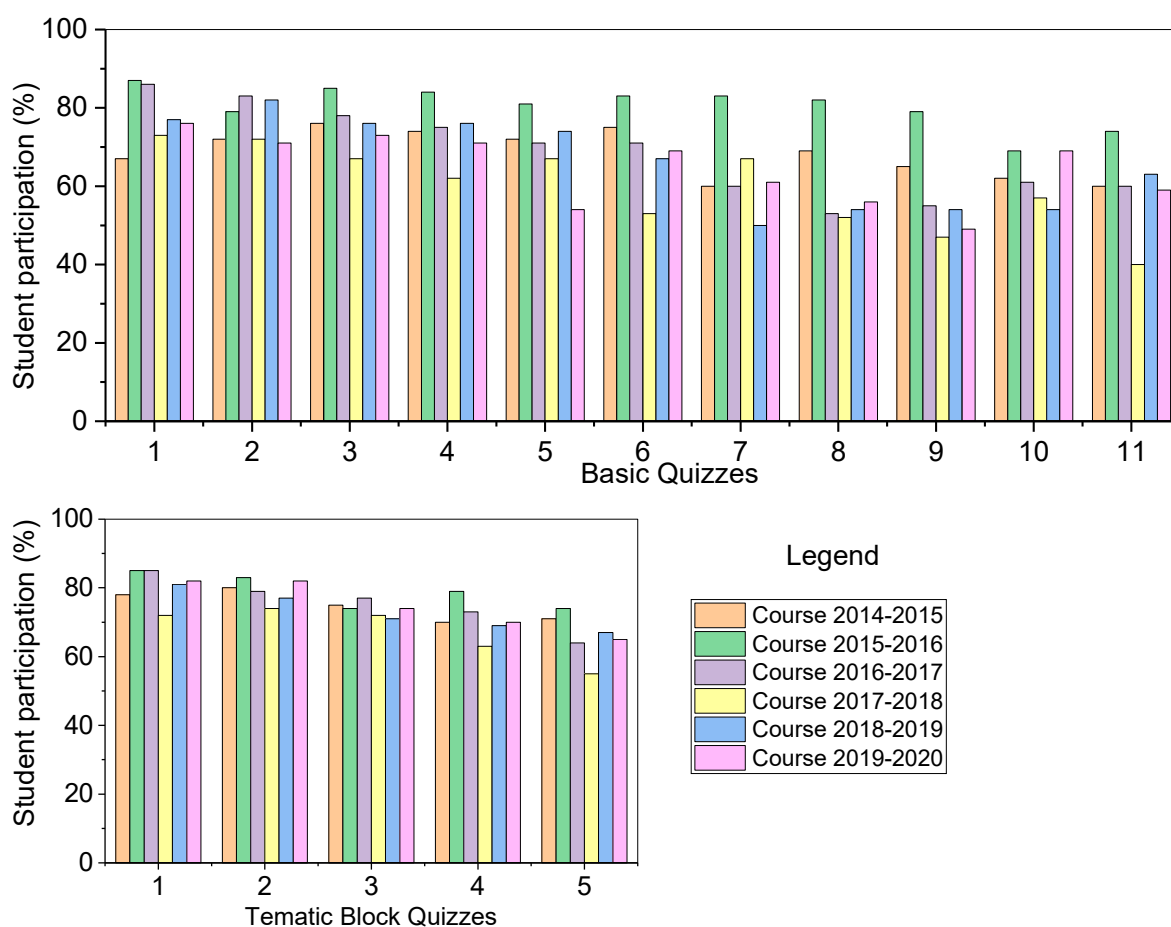
**Figure 1.** Participation in basic (BQ) and thematic block (TBQ) quizzes during the last six academic years. Own elaboration based on this study.

A detailed analysis by academic year allows us to know the dynamics and evolution of the participation. The participation falls down in the last quizzes, being always slightly lower than the first ones. This decrease is more striking in the 2016–2017 and 2017–2018 academic years, which go from approximately 85% and 75% in the first BQ to 60% and 40% in the last quiz, respectively, while it goes from 85% and 70% to 60% and 55% in the TBQs, respectively.

The general trend is a progressive decrease in participation throughout the semester in any academic year. This is due to several factors, such as possible changes in the enrolment of the subject given that some students are waiting for a possible change to another degree at the beginning of the course, and this process does not materialize until after a month, but in the meantime, they have been taking the quizzes. Moreover, mid-semester partial exams of other matters are held, so students are immersed in the study of other subjects and end up not doing the quizzes, either because the time has passed to do it, or because they have not studied. Additionally, at the end of the semester, a large number of students have decided to drop out of the degree in chemistry and are not involved in the training activities. The dropout rate in this first-degree course is approximately 20–25%. In the initial survey of the class, 25% of the students consider that the degree in chemistry is not their vocation and it was not their first option in the university pre-registration.

### 3.2. Statistical and Psychometric Data of the Quizzes and Each Item

The results provided directly by the Moodle platform (https://docs.moodle.org/dev/Quiz_statistics_calculations (accessed on 22 July 2021)) [20,21] have been analyzed and calculated according to the classical test theory [32,33]. Supplementary Materials summarize the definition of psychometric parameters. Tables 2 and 3 collect, for each quiz,

statistical data such as the average score, the standard deviation (SD), the range of correct answers (maximum and minimum percentage), also called the facility index (FI), and the asymmetry in the distribution of the scores, also called bias, together with the internal consistency coefficient (ICC), or Cronbach's alpha, which gives an idea of the quality of the tests and allows to recognize if the whole exam is homogeneous.

**Table 2.** Statistical data corresponding to the eleven basic quizzes (BQ) in different academic years.

| | 2014–2015 | | | | | 2015–2016 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **BQ** | **FI (%)** | **Average Score ± Error** | **SD (%)** | **Bias** | **ICC (%)** | **FI (%)** | **Average Score ± Error** | **SD (%)** | **Bias** | **ICC (%)** |
| 1 | 68–94 | 8.24 ± 1.20 | 24.36 | −1.29 | 78.85 | 31–95 | 6.98 ± 1.43 | 21.34 | −0.28 | 54.55 |
| 2 | 78–97 | 9.04 ± 0.80 | 16.50 | −2.90 | 76.50 | 83–97 | 9.35 ± 0.63 | 13.40 | −2.38 | 74.30 |
| 3 | 88–98 | 9.47 ± 0.65 | 10.26 | −2.20 | 59.52 | 80–100 | 9.24 ± 0.76 | 12.94 | −1.90 | 65.23 |
| 4 | 86–97 | 9.11 ± 0.79 | 15.66 | −2.38 | 74.45 | 85–100 | 9.46 ± 0.66 | 9.50 | −2.18 | 50.73 |
| 5 | 90–99 | 9.41 ± 0.61 | 14.89 | −2.99 | 82.97 | 90–98 | 9.44 ± 0.65 | 12.04 | −2.33 | 70.83 |
| 6 | 81–95 | 8.36 ± 0.78 | 21.20 | −2.50 | 86.40 | 15–97 | 8.48 ± 0.73 | 17.90 | −2.62 | 83.60 |
| 7 | 71–100 | 9.22 ± 0.59 | 10.50 | −1.66 | 67.90 | 47–97 | 8.53 ± 0.73 | 15.40 | −2.86 | 77.60 |
| 8 | 85–98 | 9.36 ± 0.66 | 13.90 | −2.46 | 77.30 | 93–97 | 9.56 ± 0.53 | 13.20 | −5.47 | 84.00 |
| 9 | 95–100 | 9.77 ± 0.45 | 6.80 | −3.38 | 56.40 | 89–98 | 9.62 ± 0.56 | 9.10 | −2.93 | 62.30 |
| 10 | 80–92 | 8.46 ± 0.85 | 20.80 | −2.08 | 83.50 | 37–90 | 6.51 ± 0.10 | 27.80 | −0.54 | 84.70 |
| 11 | 93–100 | 9.68 ± 0.48 | 10.00 | −5.42 | 76.90 | 58–100 | 8.93 ± 0.89 | 10.50 | −1.38 | 27.80 |
| | **2016–2017** | | | | | **2017–2018** | | | | |
| **BQ** | **FI (%)** | **Average Score ± Error** | **SD (%)** | **Bias** | **ICC (%)** | **FI (%)** | **Average Score ± Error** | **SD (%)** | **Bias** | **ICC (%)** |
| 1 | 81–98 | 9.35 ± 0.78 | 13.07 | −2.85 | 64.11 | 84–98 | 8.93 ± 0.88 | 19.54 | −2.07 | 79.34 |
| 2 | 85–100 | 9.54 ± 0.63 | 8.30 | −2.09 | 42.50 | 87–98 | 9.38 ± 0.63 | 14.80 | −2.78 | 81.60 |
| 3 | 97–100 | 9.86 ± 0.37 | 3.50 | −2.10 | −15.84 | 74–98 | 9.01 ± 0.85 | 14.09 | −1.53 | 62.9 |
| 4 | 77–100 | 9.25 ± 0.73 | 13.16 | −1.98 | 68.83 | 87–96 | 9.22 ± 0.75 | 14.19 | −1.83 | 71.78 |
| 5 | 91–98 | 9.63 ± 0.53 | 10.03 | −3.47 | 72.16 | 85–98 | 9.04 ± 0.78 | 17.62 | −1.78 | 80.33 |
| 6 | 52–95 | 8.48 ± 0.88 | 21.30 | −2.11 | 82.90 | 77–98 | 8.66 ± 0.82 | 23.60 | −1.96 | 87.90 |
| 7 | 62–100 | 9.11 ± 0.68 | 10.30 | −1.24 | 56.00 | 83–100 | 8.98 ± 0.71 | 16.80 | −1.76 | 82.10 |
| 8 | 86–96 | 9.10 ± 0.82 | 14.70 | −1.75 | 68.70 | 92–98 | 9.48 ± 0.64 | 11.60 | −2.20 | 70.10 |
| 9 | 82–100 | 9.35 ± 0.72 | 10.90 | −1.71 | 56.10 | 91–100 | 9.53 ± 0.62 | 10.00 | −2.49 | 61.30 |
| 10 | 51–94 | 7.73 ± 1.02 | 27.00 | −1.39 | 85.80 | 74–97 | 8.36 ± 0.79 | 25.00 | −1.90 | 90.10 |
| 11 | 88–100 | 9.55 ± 0.60 | 9.60 | −2.96 | 60.80 | 82–100 | 9.43 ± 0.68 | 10.10 | −1.53 | 54.90 |
| | **2018–2019** | | | | | **2019–2020** | | | | |
| **BQ** | **FI (%)** | **Average Score ± Error** | **SD (%)** | **Bias** | **ICC (%)** | **FI (%)** | **Average Score ± Error** | **SD (%)** | **Bias** | **ICC (%)** |
| 1 | 81–100 | 9.16 ± 0.78 | 18.01 | −2.36 | 80.85 | 95–65 | 8.59 ± 0.89 | 21.37 | −1.57 | 76.43 |
| 2 | 90–97 | 9.54 ± 0.59 | 11.10 | −2.90 | 71.50 | 99–89 | 9.34 ± 0.67 | 12.40 | −2.38 | 66.50 |
| 3 | 93–98 | 9.65 ± 0.53 | 8.88 | −2.90 | 64.26 | 93–100 | 9.46 ± 0.62 | 10.81 | −2.72 | 62.93 |
| 4 | 89–97 | 9.47 ± 0.64 | 11.37 | −2.64 | 68.22 | 86–97 | 9.26 ± 0.61 | 16.98 | −3.36 | 84.93 |
| 5 | 93–100 | 9.48 ± 0.61 | 12.41 | −2.77 | 76.22 | 91–100 | 9.52 ± 0.52 | 13.70 | −3.21 | 84.02 |
| 6 | 82–97 | 8.88 ± 0.79 | 22.40 | −2.26 | 87.60 | 75–99 | 9.05 ± 0.65 | 16.20 | −2.07 | 80.30 |
| 7 | 96–100 | 9.83 ± 0.31 | 7.50 | −5.40 | 82.90 | 85–97 | 9.25 ± 0.53 | 15.70 | −2.40 | 86.30 |
| 8 | 92–100 | 9.69 ± 0.53 | 7.50 | −3.05 | 50.50 | 93–100 | 9.79 ± 0.42 | 6.20 | −4.47 | 51.90 |
| 9 | 92–100 | 9.83 ± 0.39 | 5.00 | −3.96 | 40.50 | 84–100 | 9.45 ± 0.55 | 13.90 | −3.56 | 81.80 |
| 10 | 60–100 | 9.29 ± 0.50 | 10.80 | −4.08 | 78.70 | 70–98 | 9.05 ± 0.59 | 18.90 | −3.26 | 87.70 |
| 11 | 50–100 | 8.24 ± 0.96 | 11.90 | −0.54 | 34.90 | 93–100 | 9.90 ± 0.31 | 3.60 | −3.80 | 25.50 |

FI (Facility Index); SD (Standard Deviation); ICC (Internal consistency coefficient, Cronbach's alpha). Own elaboration based on this study.

The average score of any BQ in any academic year is high, between remarkable and outstanding (6.51 for BQ-10 and 9.83 for BQ-7 in the 2015–2016 and 2018–2019 academic years, respectively), with a high percentage of correct answers in each quiz that ranges from 70% to 100%, except in the 2015–2016 academic year where a minimum success rate of 15%, 47% and 37%, was obtained in the BQ-6, BQ-7 and BQ-10, respectively, which correspond to the two most difficult topics to assimilate: thermodynamics and electrochemistry. This large range in the correct answers yields an asymmetric distribution with a negative bias greater than −1 in all academic years. That indicates the lack of discrimination among those students who do better than the average ratio, and it is due to the fact that most of the items are classified as basic knowledge, and also to the type of question (true/false)

which shows a random response of 50%. The standard deviation is practically around 20%, except in some cases with a slightly higher value, between 22 and 28%, in those quizzes corresponding to the topics of thermodynamics and electrochemistry.

**Table 3.** Statistical data corresponding to the five thematic block quiz (TBQ) in different academic years.

| | 2014–2015 | | | | | 2015–2016 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **TBQ** | **FI (%)** | **Average Score ± Error** | **SD (%)** | **Bias** | **ICC (%)** | **FI (%)** | **Average Score ± Error** | **SD (%)** | **Bias** | **ICC (%)** |
| 1 | 62–76 | 6.93 ± 1.57 | 20.20 | −0.76 | 39.60 | 53–78 | 6.36 ± 1.66 | 17.70 | −0.49 | 12.40 |
| 2 | 52–69 | 6.08 ± 1.66 | 19.00 | −0.14 | 23.40 | 50–64 | 5.74 ± 1.68 | 21.60 | 0.17 | 39.60 |
| 3 | 48–53 | 5.16 ± 1.62 | 20.90 | −0.30 | 39.80 | 48–58 | 5.18 ± 1.59 | 21.60 | −0.09 | 45.40 |
| 4 | 34–52 | 4.48 ± 1.63 | 25.30 | −0.13 | 58.70 | 39–66 | 5.30 ± 1.64 | 22.60 | −0.24 | 47.50 |
| 5 | 28–51 | 3.93 ± 1.70 | 21.90 | 0.18 | 40.00 | 38–54 | 4.70 ± 1.68 | 24.30 | −0.37 | 52.20 |
| | **2016–2017** | | | | | **2017–2018** | | | | |
| **TBQ** | **FI (%)** | **Average Score ± Error** | **SD (%)** | **Bias** | **ICC (%)** | **FI (%)** | **Average Score ± Error** | **SD (%)** | **Bias** | **ICC (%)** |
| 1 | 61–81 | 6.97 ± 1.55 | 19.80 | −0.58 | 38.60 | 62–74 | 7.01 ± 1.52 | 21.60 | −0.78 | 50.40 |
| 2 | 61–76 | 5.80 ± 1.57 | 18.50 | −0.72 | 28.20 | 50–64 | 5.23 ± 1.72 | 21.50 | −0.66 | 35.90 |
| 3 | 40–60 | 5.23 ± 1.62 | 22.90 | −0.03 | 50.00 | 42–61 | 5.14 ± 1.62 | 23.70 | −0.07 | 53.10 |
| 4 | 48–66 | 5.86 ± 1.68 | 22.40 | −0.30 | 43.80 | 46–76 | 5.82 ± 1.65 | 21.40 | −0.73 | 40.40 |
| 5 | 42–64 | 5.68 ± 1.67 | 19.40 | −0.35 | 26.50 | 50–73 | 5.88 ± 1.57 | 23.10 | −0.33 | 54.10 |
| | **2018–2019** | | | | | **2019–2020** | | | | |
| **TBQ** | **FI (%)** | **Average Score ± Error** | **SD (%)** | **Bias** | **ICC (%)** | **FI (%)** | **Average Score ± Error** | **SD (%)** | **Bias** | **ICC (%)** |
| 1 | 55–78 | 6.81 ± 1.57 | 19.50 | −0.64 | 35.50 | 61–75 | 6.99 ± 1.09 | 19.4 | −1.02 | 35.40 |
| 2 | 51–70 | 6.24 ± 1.65 | 20.60 | −0.50 | 35.20 | 50–68 | 6.00 ± 1.00 | 21.0 | −0.28 | 38.40 |
| 3 | 43–58 | 5.15 ± 1.67 | 19.40 | 0.07 | 25.70 | 49–57 | 5.31 ± 0.86 | 22.3 | −0.51 | 46.50 |
| 4 | 41–62 | 5.32 ± 1.76 | 16.80 | 0.03 | −10.30 | 49–65 | 5.73 ± 0.93 | 19.9 | −0.07 | 32.30 |
| 5 | 44–56 | 4.96 ± 1.69 | 18.70 | 0.02 | 18.30 | 43–56 | 4.95 ± 0.81 | 20.6 | −0.48 | 36.50 |

FI (Facility Index); SD (Standard Deviation); ICC (Internal consistency coefficient, Cronbach's alpha). Own elaboration based on this study.

The ICC in most quizzes at any academic year is higher than 65%, the minimum value proposed as indicator of an overall homogeneity of the quiz [34]. However, in some cases, values lower than 65% have been obtained, for example in the BQ-3 and BQ-9 in the 2014–2015 academic year, even significantly lower values of 27.80% in the BQ-11 of the 2015–2016 academic year or even negative values of −15.84% in the BQ-3 of the 2016–2017 academic year. These results show the limitation of this parameter in considering that the quiz measures with the same precision all the students evaluated when it really depends on the level of each student and, ultimately, on the population used to calculate it.

Moreover, the dispersion of the IF and the discrimination index (DI) for each item of any quiz have been analyzed in order to know the item effectiveness to discern between students with different cognitive ability (Figure S1, left). Most of the questions have an adequate discrimination, with a DI above 30%. A more detailed analysis of the discriminative efficiency (DE) for any item of the different BQ (Figure S2) shows that the effectiveness of the items depends on the academic year and, therefore, on the student population. For example, in the academic year 2015–2016, all items of the BQ1 do not reach 30% of DE, while in the 2018–2019 academic year, they are all above 30%. The same behavior has been found in other items corresponding to other quizzes. Therefore, the same item may or may not be discriminatory for a population of students depending on the level of knowledge they have, and therefore, questions that have a low DI should not be discarded. It is concluded that quizzes made only with true/false items serve as continuous training activities in the teaching–learning process of a matter, not being feasible as assessment activities because they are not discriminatory for students.

Different results are obtained for the TBQs (Table 3). The average score drops significantly with respect to the BQ quizzes, from 5.16 in the TBQ-3 of the 2014–2015 academic year to 7.01 in the TBQ-1 of the academic year 2017–2018, even in some cases reaching

values of 3.9 or 4.5 in the TBQ-4 and CBT-5 of the 2014–2015 academic year, for example. Moreover, the FI index drops significantly and ranges from 28% in the TBQ-5 of the 2014–2015 academic year to 81% in the TBQ-1 of the 2016–2017 academic year, but in no case does it reach 100% in any of the items in any quiz. The dispersion in the average scores oscillates around 20%, being slightly high in the last three CBTs of certain academic courses. The asymmetry of the distribution in the scores, the so-called bias, is still negative, but now with a value lower than $-1$, reaching a slightly positive value and with an almost symmetric distribution, with a bias close to zero in the last three quizzes of the 2018–2019 academic year with a value between 0.07 and 0.02.

As a general trend, the bias in the first two quizzes, TBQ-1 and TBQ-2, is somewhat higher than in the rest of the TBQ-3, TBQ-4 and TBQ-5 ones that have a bias close to zero. This indicates that the topics corresponding to the first two blocks are better assimilated than the rest of the topics corresponding to the blocks of thermodynamics, electrochemistry and kinetics. This effect is probably due to the fact that the first topics are already studied in the bachelor grade, while the topics of the last blocks are totally new, which means an effort in learning process.

Therefore, these TBQs are more discriminative between students than BQs, that is, here the cognitive abilities of each student are tested. The FI-DI scatter diagrams are shown in Figure S1 right. Although a low DI value is obtained due to the random characteristic of the quiz, the detailed analysis of the FI-DI diagrams for all items in any TBQ (Figure S3) reveals that most of items are discriminative with a DI value above 30% and with a wide range in the FI. However, the ICC is always lower than the reference value of 65% [34]. This is because the items of the quiz have been randomly selected by the Moodle platform and show different questions for each student.

### 3.3. Students' Opinion on the Educational Activities

A survey was performed among the students to find out their opinion on these two types of virtual quizzes. It is a short survey with only eight statements, four for each type of quizzes, in which students have to answer YES or NO to the proposed statement. Although this questionnaire was not contrasted by the scientific community and its experimental validity was not demonstrated, it only tries to show the students' opinion on Moodle quiz activities in few ideas. Table 4 shows the average values obtained in the BQ and TBQ, respectively, during the last six academic years. These average values practically coincide with those obtained for each academic course; therefore, the students' opinion about how they apperceive the virtual quizzes, practically, does not change with the years.

**Table 4.** Students' opinion on the BQ and TBQ. Average results of the last six academic years. Own elaboration based on this study.

| | Statements | Responses (Average Values) | | |
|---|---|---|---|---|
| | **Basic Quizzes (BQ)** | **YES** | **NO** | **NK/NA** |
| 1.- | The difficulty of the question in these quizzes has been adequate for a basic level | 89.9% | 5.4% | 4.7% |
| 2.- | The time limit of one hour is sufficient to take the quiz | 89.2% | 5.4% | 5.4% |
| 3.- | The quizzes allow you to study the subject continuously | 59.4% | 32.4% | 8.2% |
| 4.- | The quizzes are used for self-evaluation of the subject and know the level of knowledge acquired | 67.6% | 27.1% | 5.3% |
| | **Statements** | **Responses (Average Values)** | | |
| | **Thematic Block Quizzes (TBQ)** | **YES** | **NO** | **NK/NA** |
| 1.- | The level of difficulty of these quizzes is higher than that of the basic ones | 91.9% | 5.4% | 2.7% |
| 2.- | The time limit of one hour is sufficient to take the quiz | 51.3% | 45.9% | 2.8% |
| 3.- | These quizzes carried out on dates programmed allow you to study the topics before evaluation | 62.1% | 35.1% | 2.8% |
| 4.- | These quizzes serve to strengthen the acquired knowledge | 59.4% | 37.8% | 2.8% |

Regarding BQ, most of the students, about 90%, tell that the difficulty of the items is adequate for a basic level and the time of one hour is more than enough. These BQs help a lower percentage 60% to study the matter continuously, while the rest, 30%, think that they do not study continuously, even though they have to do the online quizzes. However, the BQs serve as a self-assessment of the knowledge acquired in class for almost 70%. Only 5–8% of students do not know or do not answer (NK/NA) to each of the proposed statements. The results obtained for the TBQs are a little bit different.

Regarding the TBQ, practically all students, 92%, say that the level of difficulty in these quizzes is greater than that of the BQ. This was predictable because one of the goals of these TBQ was to measure the level of abstract reasoning and assimilation of theoretical concepts. Only 50% of the students say that time limit of one hour is sufficient. On the other hand, these quizzes scheduled in fixed days allow more than 60% of students to study the days before doing the test. This fact consolidates the knowledge acquired in the classroom. Only 3% of the students answer NK/NA to the proposed statements.

In short, both types of quiz fulfill the goal of favoring a continuous study of the matter along the semester, avoiding the general trend to study only the days before the written exam. Moreover, all these educational activities into an electronic environment promote an online-accessible medium for the student that serves as a self-assessment of the level of knowledge acquired.

*3.4. Final Scores in General Physical Chemistry*

Figure 2 shows the final scores in the matter of general physical chemistry corresponding to the two Ordinary Calls of exams (February and September), between 2011 to 2014 courses following a traditional teaching methodology, and between 2014 to 2020 courses following the new methodology with quizzes.

In the first Ordinary Call (February), a slight decrease in the percentage of students not presented and also in that of failed following the new methodology is observed. However, it is interesting to note the increase in the number of passes, not only with the minimum score but also with remarkable, outstanding and even honors (H) in the last five academic years, highlighting the academic year 2016–2017 where the percentage of outstanding is higher than in the rest of the courses.

In the second Ordinary Call (September), the percentage of students not presented remains almost constant in all the courses developed with the methodology based on quizzes and is practically the same as that of the first courses taught with the traditional methodology. There is not much variation in the percentage of failures and approved following one methodology or another, except in the last academic year 2019–2020, in which the number of not presented decreases while the number of approved increases. In this September call, there is a low percentage of outstanding students and the absence of scores above the remarkable following the new methodology with quizzes.

These online education activities help not only students with high cognitive capacity to attain good scores but also those with a medium level to pass the exam in the first call of February. The percentage of students that remain for the second call in September are really those who have not assimilated the physical chemistry knowledge and those who find it difficult to make scientific reasoning or deductions.
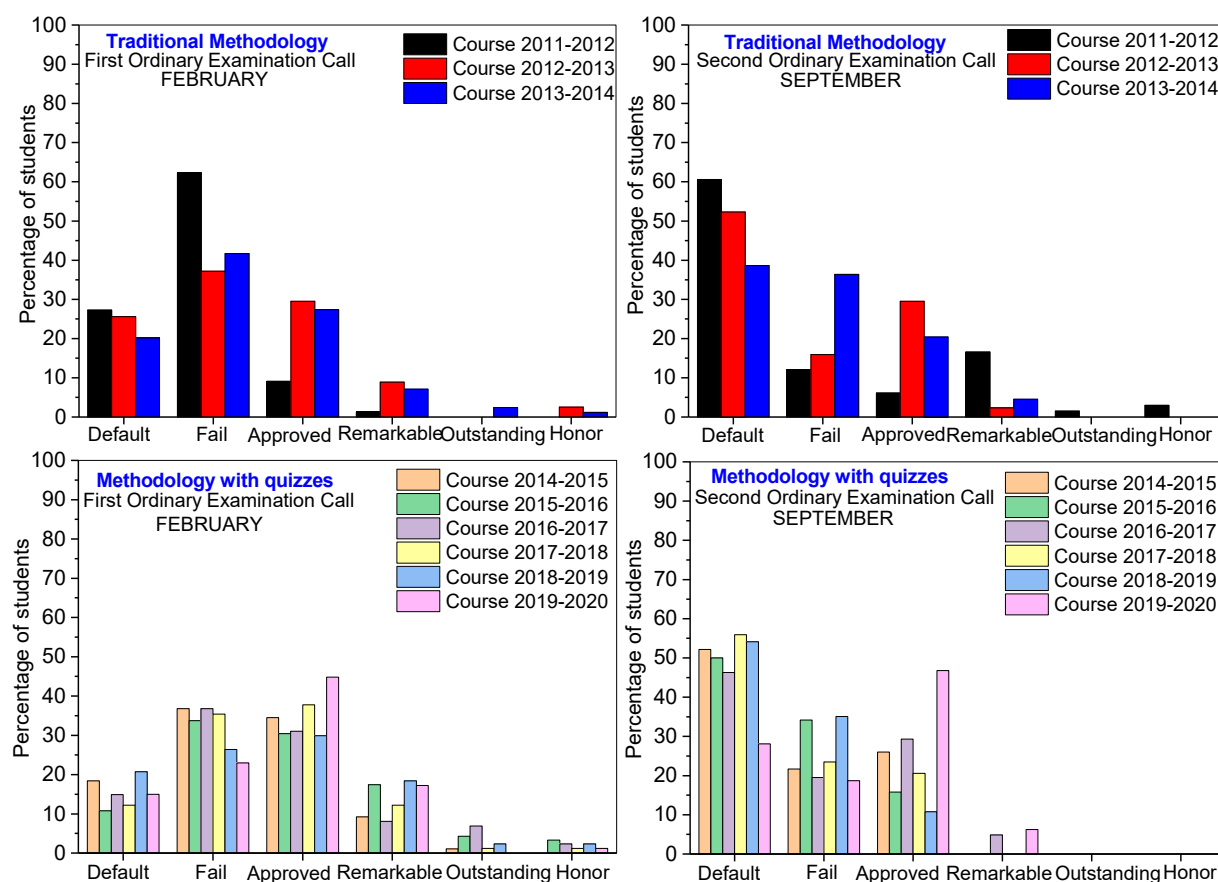
**Figure 2.** Scores of General Physical Chemistry in Ordinary Calls of exams. Own elaboration based on this study.

## 4. Conclusions

The use of Moodle quizzes, as online activities, favors the implementation of a different educational methodology in the subject of general physical chemistry of the Degree in Chemistry. However, not all quizzes can be used as assessment tools given that the item type and the quiz settings play an important role. In this work, two different types of quiz are proposed, basic (BQ) and thematic block (TBQ). BQ has only true/false items and these are the same for all students, while TBQ has multiple choice, matching and numerical items that are randomly selected from a category of the question bank.

Statistical and psychometric data provided by Moodle platform were analyzed. Most of items show a discrimination (DI) and facility (FI) index according to the proposed Moodle reference range. The FI-DI dispersion graphs, together with the average scores and the bias values, show the different quality of the two types of quiz. The BQs can be used as formative teaching activities because they have not enough evaluative quality to distinguish the different capacities and abilities of students, yielding in any quiz higher average scores and FI values, and also a large negative bias. TBQs are more discriminative, showing lower values of both FI and average scores, with a bias near to zero value, and thus, these can discern competencies and skills among students, so they could be used as assessment tools. As a conclusion, the true/false item must not be used as evaluative item. In the future, it is necessary to study how each type of item (multiple choice, numerical and matching) contributes individually to the final scores of the quiz in a similar performance. This would allow us to select the best type of item to use it in a particular assessment quiz.

Although the methodology applied is weak, the performance developed along the years indicates that TBQs work quite well to evaluate the physical chemistry knowledge and the different capacities of students, independent of the student population, and it could even be extrapolated to other similar scientific scenes.

The analysis of statistic data in TBQs, like FI index, allow the teacher to know the topics that were not well understood by the students and gives a feedback on learning process of the students. Moreover, these quizzes work as a self-assessment for the students, providing them a better preparation for the written exam. The performance of these quizzes has not supposed an excessive workload for the students, allowing improved scores obtained in the first Ordinary Call of February, unlike those obtained following a traditional methodology in previous courses.

In few words, the Moodle statistics point out that a particular quiz has assessment quality if the bias is symmetric, around zero value, and the range of FI and DI is about 40–70%. This feature can be obtained using a huge question bank with the aim to perform a random quiz. Mixed items (multiple choice, numerical and matching) also contribute to achieving a quiz in which different skills and abilities of physical chemistry have been inquired. In addition, the similar results obtained in this experience along the years, with different students' population, would also prove the validity and reliability of the designed quizzes. This study shows how the analysis of statistical and psychometric parameters allows checking if the design of an educative activity based on Moodle quizzes can be used as an assessment tool to evaluate the skill and competences of a particular subject along the semester.

## References

1. Valverde-Berrocoso, J.; Garrido-Arroyo, M.C.; Burgos-Videla, C.; Morales-Cevallos, M.B. Trends in Educational Research about e-Learning: A Systematic Literature Review (2009–2018). *Sustainability* **2020**, *12*, 5153. [CrossRef]
2. Linuma, M. *Learning and Teaching with Technology in the Knowledge Society*; Springer: Singapore, 2016.
3. Luisa Sevillano-García, M.A.; Vázquez-Cano, E. The Impact of Digital Mobile Devices in Higher Education. *J. Educ. Technol. Soc.* **2015**, *18*, 106–118. Available online: http://www.jstor.org/stable/jeductechsoci.18.1.106 (accessed on 10 March 2021).
4. Delialioglu, O.; Yildirim, Z. Design and development of a technology enhanced hybrid instruction bases on MOLTA model: Its effectiveness in comparison to traditional instruction. *Comput. Educ.* **2008**, *51*, 474–483. [CrossRef]
5. Ryberg, T.; Niemczik, C.; Brenstein, E. Methopedia-Pedagogical Design for European Educators. In Proceedings of the 8th European Conference on e-Learning, Bari, Italy, 29–30 October 2009; pp. 503–511.
6. Tamin, R.M.; Bernard, R.M.; Borokhovski, E.; Abrami, P.C.; Schmid, R.F. What Forty Years of Research Says About the Impact of Technology on Learning: A Second-Order Meta-Analysis and Validation Study. *Rev. Educ. Res.* **2011**, *81*, 4–28. [CrossRef]
7. Francis, R.; Shannon, S.J. Engaging with blended learning to improve students' learning outcomes. *Eur. J. Eng. Educ.* **2013**, *38*, 1–11. [CrossRef]
8. Kesici, S.; Sahin, I.; Akturk, A.O. Analysis of cognitive learning strategies and computer attitudes, according to college students' gender and locus of control. *Comput. Hum. Behav.* **2009**, *25*, 529–534. [CrossRef]
9. López-Pérez, M.V.; Pérez-López, M.C.; Rodríguez-Ariza, L. Blended learning in higher education: Students' perceptions and their relation to outcomes. *Comput. Educ.* **2011**, *56*, 818–826. [CrossRef]
10. Randy Garrison, D.; Vaughan, N.D. *Blended Learning in Higher Education: Framework, Principles, and Guidelines*; John Wiley & Sons, Inc.: New York, NY, USA, 2008; ISBN 978-1-118-26955-8.

11. Anthony, B.; Kamaludin, A.; Romli, A.; Raffei, A.F.M.; Phon, D.N.; Abdullah, A.; Ming, G.L. Blended Learning Adoption and Implementation in Higher Education: A Theoretical and Systematic Review. *Technol. Knowl. Learn.* **2020**, *25*, 1–48. [CrossRef]

12. Bidarra, J.; Rusman, E. Towards a pedagogical model for science education: Bridging educational contexts through a blended learning approach. *Open Learn. J. Open Distance E-Learn.* **2017**, *32*, 6–20. [CrossRef]

13. Jamshidifarsani, H.; Tamayo-Serrano, P.; Garbaya, S.; Lim, T. A three-step model for the gamification of training and automaticity acquisition. *J. Comput. Assist. Learn.* **2021**, *37*, 994–1014. [CrossRef]

14. Kirzner, R.S.; Alter, T.; Hughes, C.A. Online Quiz as Exit Ticket: Using Technology to Reinforce Learning in Face to Face Classes. *J. Teach. Soc. Work* **2021**, *41*, 151–171. [CrossRef]

15. Manzano-Leon, A.; Camacho-Lazarraga, P.; Guerrero, M.A.; Guerrero-Puerta, L.; Aguilar-Parra, J.M.; Trigueros, R.; Alias, A. Between Level Up and Game Over: A Systematic Literature Review of Gamification in Education. *Sustainability* **2021**, *13*, 2247. [CrossRef]

16. Lo, C.; Tang, K. Blended Learning with Multimedia e-Learning in Organic Chemistry Course. In Proceedings of the 2018 International Symposium on Educational Technology (ISET), Osaka, Japan, 31 July–2 August 2018; pp. 23–25. [CrossRef]

17. Nash, S.S.; Rice, W. *Moodle 3 E-Learning Course Development*; PACKT Publisher: Birmingham, UK, 2015.

18. Brophy, J.; Alleman, J. Assessment in a Social Constructivist Classroom. *Soc. Educ.* **1998**, *62*, 32–34.

19. Gold, S. Constructivist approach to online training for online teachers. *J. Asynchronous Learn. Netw.* **2001**, *5*, 35–57. Available online: https://wikieducator.org/images/f/fb/ALN_Constructivist_Approach.pdf (accessed on 2 March 2021). [CrossRef]

20. Gamage, S.H.; Ayres, J.R.; Behrend, M.B.; Smith, E.J. Optimising Moodle quizzes for online assessment. *Int. J. STEM Educ.* **2019**, *6*, 1–14. [CrossRef]

21. Gómez-Soberón, J.; Gómez-Soberón, M.C.; Corral-Higuera, R.; Arredonde-Rea, S.P.; Almaral-Sánchez, J.L.; Cabrera-Cavarrubias, F.G. *Calibrating Questionnaires by Psychometric Analysis to Evaluate Knowledge*; SAGE: Thousand Oaks, CA, USA, 2013. [CrossRef]

22. Jaeger, M.; Adair, D. Time pressure in scenario-based online construction safety quizzes and its effect on students' performance. *Eur. J. Eng. Educ.* **2017**, *42*, 241–251. [CrossRef]

23. Krause, C.M.A.; Krause, R.B.S.; Krause, R.B.S.; Gomez, N.G.M.D.; Jafry, Z.M.D.; AmDinh, V. Effectiveness of a 1-Hour Extended Focused Assessment with Sonography in Trauma Session in the Medical Student Surgery Clerkship. *J. Surg. Educ.* **2017**, *74*, 968–974. [CrossRef] [PubMed]

24. Sullivan, D.P. An Integrated Approach to Preempt Cheating on Asynchronous, Objective, Online Assessments in Graduate Business Classes. *Online Learn.* **2016**, *20*, 195–209. Available online: http://onlinelearningconsortium.org/read/online-learning-journal/ (accessed on 4 May 2021). [CrossRef]

25. Burden, P. ELT teacher views on the appropriateness for teacher development of end of semester student evaluation of teaching in a Japanese context. *System* **2008**, *36*, 478–491. [CrossRef]

26. Nulty, D.D. The adequacy of response rates to online and paper surveys: What can be done? *Assess. Eval. High. Educ.* **2008**, *33*, 301–314. [CrossRef]

27. Shepard, L.A. *The Assessment in the Classroom: Texts Evaluation*; National Institute for Educational Evaluation: Distrito Federal, Mexico, 2006.

28. Ferrao, M. E-assessment within the Bologna paradigm: Evidence from Portugal. *Assess. Eval. High. Educ.* **2010**, *35*, 819–830. [CrossRef]

29. Blanco, M.; Ginovart, M. Moodle quizzes for assessing statistical topics in engineering studies. In Proceedings of the Joint International IGIP-SEFI Annual Conference 2010, Trnava, Slovakia, 19–22 September 2010. Available online: https://upcommons.upc.edu/bitstream/handle/2117/9992/blanco_ginovart_igip_sefi_2010.pdf (accessed on 9 August 2021).

30. Blanco, M.; Ginovart, M. On How Moodle Quizzes Can Contribute to the Formative e-Assessment of First-Year Engineering Students in Mathematics Courses. *RUSC Univ. Knowl. Soc. J.* **2012**, *9*, 166–183. [CrossRef]

31. Crocker, L.; Algina, J. *Introducction to Classical and Modern Test Theory*; Holt Rinehart and Winston: New York, NY, USA, 1986.

32. Wang, Z.; Osterlind, S.J. Classical Test Theory. In *Handbook of Quantitative Methods for Educational Research*; Teo, T., Ed.; Sense Publishers: Rotterdam, The Netherlands, 2013. [CrossRef]

33. Shrout, P.E.; Lane, S.P. Psychometrics. In *Handbook of Research Methods for Studying Daily Life*; Mehl, M.R., Conner, T.S., Eds.; The Guilford Press: New York, NY, USA, 2012; pp. 302–320.

34. GNU. General Public License. 2013. Available online: https://docs.moodle.org/dev/Quiz_report_statistic (accessed on 7 July 2020).