

# The Similarity Index: A New and Simple Algorithm to Measure Coherence in Polls

## —A Case Study Where the Students Evaluate Assessments in a General Chemistry Course

Patricio Flores-Morales

Departamento de Polímeros, Universidad de Concepción, Concepción, Chile

Email: patflores@udec.cl

**How to cite this paper:** Flores-Morales, P. (2022). The Similarity Index: A New and Simple Algorithm to Measure Coherence in Polls. *Creative Education*, 13, 203-237. <https://doi.org/10.4236/ce.2022.131014>

**Received:** November 24, 2021

**Accepted:** January 22, 2022

**Published:** January 25, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

### Abstract

Assessments have been a crucial part of the process of learning and teaching. Most of the time, this process has been measured with summative assessments but what about formative ones? Historically, the teacher chooses the appropriate assessment for his/her class. However, what do the students have to say about that? In this case study, we present the results of a Likert-type poll for five assessments applied in a General Chemistry course. Students gave their answers according to three measurable intended aspects: the cognitive, the emotional and the social. To measure the coherence of those three aspects a similarity index was created. This instrument, together with the Likert-type poll, allowed us to advise which assessment should be accepted or rejected according to the student's point of view. From a variety of assessments, the student selected the following increasing ranking of assessments: Short Multiple-Choice Questions, General Questions/Answers Out Loud and Prediction-Observation-Explanation.

### Keywords

Student Evaluation, Evaluation Methods, Formative Assessment, Similarity Index, First-Year Undergraduate, Chemical Education Research

---

## 1. Introduction

The social media (SM) allow users to evaluate their peers by pressing “like”. Young people are the common users of SM, so they are habituated to be evaluated and evaluating. Even when this is a spread practice among the Z-generation, it seems such a behaviour is not extrapolated to assessments.

A discussion on how the students should be assessed has had a significant impact over the last years (Creme, 2005; Maxwell, 2012; Mayowski et al., 2018; Schneider et al., 2019; Smith, 2007; Stewart & Richardson, 2000). Should it be at the beginning, in the middle or at the end of a unit? It seems that a general conclusion applies; assessment must be a journey, not an objective (Bulwik, 2004; Chamizo, 1995; Viera et al., 2007).

Chamizo (1995, 1996) argues that if the questions that are asked on a test are for specific contents, the students will rapidly forget those contents due to there are no circumstances to overcome. In this sense, students must be challenged to think and reflect on why they are doing what they are doing. Thus, the assessment must be an integral process, not only a goal by itself.

Bulwik (2004) goes beyond Chamizo's ideas indicating that:

Most of the evaluation methods currently applied for learning are conservative, considering that purposes and teaching methodologies have changed.

According to Bulwik,

[...] to privilege the social character of the evaluation of learning over its pedagogical nature, led to equalise assessment, examination, and qualification.

Moreover, she tops off saying:

[*the previous statement led*] to consider that assessments are only for rating an exam, turning the exams and tests in moments of tension as much the teacher as the students.

The construction of knowledge is dynamic; the meanings are revisited and updated, recurrently, so the student must do his/her own learning pathway (Price et al., 2017). Thus, if the students take part in the assessment, the commitment to learn is more significant (Bulwik, 2004). A shortlist of formative/summative assessments is presented in **Appendix A** (view supporting information, from now on S. I.).

Since the assessment is a trip and not a goal, it seems natural questioning: what are the advisable assessments to use? According to literature (Hortigüela et al., 2019; Prashanti & Ramnarayan, 2019; Quesada et al., 2019), this is a critical issue when deep/active/lifelong learning (Davari & Bahraman, 2019; Lewis, 2019; Rubenson, 2019) is the objective needed for reaching.

The main issue with assessments is it is the teacher who makes these instruments. He/she must have the expertise not only in his/her field, but also in the design of assessments. This issue depends on his/her career, so some teachers have the assessments design skills more developed than others. But, in what choices the assessments design is based? (Fischman et al., 2019; Tan, 2019). Is there any chance to involve the students asking for their opinions? We think this is an aspect that we should consider more often.

Furthermore, assessing implies not only a cognitive aspect, but an emotional and, sometimes, a social aspect, as well. When an assessment is placed, the students' mind tries to find the answer. However, a part of the students' feelings is

also involved. They can react negatively to a particular question if they are not comfortable with the content. Moreover, they can react negatively to a specific type of assessment if they do not like it. As in social media, feelings are involved when teachers assess students.

On the other hand, socio-cognition is another aspect present in assessments, even more, if the students work on groups, and they have peer review. However, the social issue is also present when students are asked if some assessment should be applied to future freshmen. In this scenario, a freshman may think not only in his/her benefits but in the repercussion for their new classmates (next year), as well. In that regard, a parameter that could simultaneously measure the coherence among the cognition, the emotional and the social aspects of the assessment could bring attention to which ones are preferred by the students.

Giving the full range of assessments listed, this study had the purpose to analyse which assessments are preferred the most by the students in a General Chemistry course.

To reach that goal, this manuscript intent to bring some light to the following question: is there any parameter to evaluate the students' choices with respect the way they are assessed?

To respond that question, a poll was conducted. Most of the surveys only evaluate the emotional part, but in this manuscript, the cognitive and social aspects were also evaluated. For that, the paper presents a parameter called "the similarity index", to simultaneously measure the cognitive, the emotional, and the social aspects to quantify the students' preferences. Thus, along this manuscript, a detailed argumentation is presented to answer the proposed question.

The manuscript is structured as follows: after this introduction, a description of the methods is presented, followed by the results and the discussion, ending with some concluding remarks.

## 2. Methods

### 2.1. General Overview

This research was made in a group of tertiary students in an anonymous way and presents a one-year case study in a General Chemistry course where, traditionally, the students were no submitted to any formative assessment. The teacher in charge was trained in general formative assessment with no specificity for Chemistry. Thus, it was the mission of the teacher to explore the most suitable assessments for this course. To accomplish that three to five formative assessment were tested in any lesson along the year (see **Table B1** and **Table B2**, **Appendix B**, S. I.).

To obtain the students' opinions concerning assessment, they were polled (by a Likert-type poll), after the finishing of every term. Moreover, a sample of them was interviewed. To gather the Likert-type poll information in a "simple" way, a similarity index was created. This index intended to measure the coherence among the cognitive, the emotional and the social aspect based on the three

questions given in the poll (see later). Thus, a high correlation (or coherence) of these three variables indicated the assessment that students appreciated the most, and a low correlation indicated the opposite thing.

According to the previous paragraph, and for the sake of clarity, the following section is ordered as follows: first, the type of lessons and the type of students is described, second, a brief review of each used assessment is given, third, the Likert-type poll details are presented together with segments of the interviews, and forth, the similarity index is presented.

## 2.2. Type of Lessons

Lectures were scheduled, as shown in **Table B1** and **Table B2** (see **Appendix B**, S. I.). Usually, a lesson consisted of an opening to the theme, an introductory activity, a development by exposition, and a closing involving some assessment listed in **Table B1** and **Table B2** (see **Appendix B**, S. I.). The opening of the lessons had the purpose of running the students' minds. Nonetheless, not always was related to students' assessment. For that, the teacher's exposition delivered the required contents to complete the assessment.

## 2.3. Type of Students

The type of the students is described primarily for General Chemistry 1 (GM1, the first-term course), and secondly for General Chemistry 2 (GM2, the second-term course).

### First-term (GM1)

The course consisted of 65 students, but since the attendance to classes is not mandatory at the University of Concepción, so an average of 60% of them attended the classes. The students that responded to the poll (see below) were 38. The course included 46% of men and 54% of women. Ages comprehended between 18 and 20 years old. The students' origin was from Concepción city and nearby cities. This origin assured a composite sample concerning the socio-ethnic origins, the regime, and the quality of education.

### Second-term (GM2)

Eighty-four students signed up in GM2. Thus, 19 students were attending the course for a second time. An average of 66% of the students participated in the classes, meaning 78 students responded to the poll (see below). The course included 42% of men and 58% of women with ages between 18 and 20 years old. The origins of the sample described in the first-term section are also valid here.

## 2.4. Formative Assessments

The students did not mandatorily answer formative assessments. Because of this, all the plots presented in this study were normalised. **Table B1** (see **Appendix B**, S. I.) gathers a list of activities, lesson-by-lesson. Recalling the list shown in **Appendix A** (S. I.), the formative assessments used in classes were one-minute pa-

per (OMP), short multiple-choice questions (SMCQ), directed paraphrasing (DP), Predict, Observe and Explain (POE), and general question/answer out loud (GQAOL). As was previously settled, the choice for these assessments was completely exploratory based on the following criteria made by the teacher in charge:

- OMP because it is an anonymous form to obtain students thoughts
- SMCQ because the students are used to work with them since high school
- DP because it is a challenge (the students never used it before)
- POE to test the students' prediction ability
- GQAOL due to it is a classical and a fast way to receive feedback from the students.

#### **One-Minute Paper**

To follow the students' progress, they answered two questions, in a piece of paper, at the end of some lessons (see **Table B1, Appendix B, S. I.**). The OMP assessment typically takes one minute (Angelo & Cross, 1993; Harwood, 1996; Kumar et al., 2017). The questions were:

- What did you like/learn most about the class?
- What content did you not understand?

The teacher analysed the answers and collected them into two groups: the ones they liked/learned the most and the ones they did not understand. This assessment allowed the teacher to understand what students "like" the most or learned and reinforce what was not clear (Kumar et al., 2017; Lutterodt, 2017; Stead, 2005). The next lesson the teacher gave feedback to the students helping them to realise their strengths and their weaknesses.

#### **Short Multiple-Choice Questions**

(Short) multiple-choice questions (SMCQ) are another form to follow the students' progress (Bresnock et al., 1989; Butler, 2018; Dodd & Leal, 1988). Three SMCQ (each of them having four choices) were asked every time this evaluation was used (at the end of the lessons, see **Table B1, Appendix B, S. I.**). The questions tuned with the contents of the corresponding class, and the students received instant feedback through the platform socrative.com (Arriaga et al., 2017; Blackburn, 2015; Frías, Arce, & Flores-Morales, 2016; Guarascio et al., 2017; Kokina & Juras, 2017; Manning et al., 2017).

#### **Directed Paraphrasing**

In directed paraphrasing (DP) (Angelo & Cross, 1993), students must explain, in their own words, a complex concept to a regular audience. Due to the students must summarise content in their own words, they prove they have learned it (Cheung, 2016; Moran et al., 2014; Tan, 2017; Uemlianin, 2000).

As an example, for this assessment, after a class of colligative properties (class 36, **Table B1, Appendix B, S. I.**), students had to demonstrate why a hand cream based on sodium chloride was better for keeping hands warm in cold days than a hand cream based on glycerol. The regular audience was their mothers (not having the chance to talk to them).

### Predict, Observe, and Explain

Predict, Observe and Explain (POE) is another formative learning method (Güven, 2014; Hong et al., 2014; Hsu et al., 2011; Kose & Bilen, 2012; White & Gunstone, 1992). In the present case, the students had to predict what was going to happen in a particular experiment, observe the chemical reaction and finally, explain in their own words what they watched.

For example, for the lesson 39 (see **Table B2, Appendix B, S. I.**), a range of acid-base chemical reactions was made as follows: a solution of sodium hydroxide  $0.05 \text{ mol}\cdot\text{L}^{-1}$  was poured into seven beakers with water and drops of different acid-base indicators. The visible spectrum range of colours appears as the water was poured. Finally, all the beakers were poured into a 1 L beaker containing 10 mL of sulfuric acid  $2.5 \text{ mol}\cdot\text{L}^{-1}$  for discolouring the solutions. At no time, the students knew what the solutions were.

The students had to follow the POE method: predict what was going to happen when the solutions were mixed, observe the colour appearance, and explain why this happened. Next, the students had to predict what was going to happen when the seven beakers were poured into the 1 L beaker containing 10 mL of sulfuric acid  $2.5 \text{ mol/L}$ , observe the discolouring and explain why.

### General Question/Answer Out Loud

Asking students general questions out loud (GQAOL) is a classical form of formative evaluation (Bannert & Mengelkamp, 2008; Lancaster, 2007; Metcalfe & Xu, 2018; Wardrop, 2012). During almost every lesson this assessment was used.

## 2.5. Poll

The poll consisted of students' perceptions about the activities. Their answers were anonymously, and not mandatory. Because of that, all the plots presented in this study, regarding the number of students, were normalised.

The students were asked three questions on the Likert-type scale modality (Likert, 1932; Matas, 2018; Orozco et al., 2018; Robertson, 2012). **Table 1** depicts the question and the measured aspect.

For this poll, the Likert-type scale meaning was: 1, nothing; 2, little; 3, enough; 4, much; 5, very much.

**Table 1.** Questions and measured aspects for the Likert-type scale asked to the students.

Question number	Question	Measured aspect
Q1	How much do you think these types of evaluations helped you to understand the contents?	Cognitive
Q2	How much would you recommend using the following assessment for General Chemistry freshmen students?	Social
Q3	How much did you like the following evaluations?	Emotional

## 2.6. Interviews

A sample of three students (two men and one woman) was interviewed in a semi-structured interview regarding their experiences and opinions of the assessments. Verbal informed consent was obtained prior to the interviews.

## 2.7. The Similarity Index

The coherence between the cognitive, the emotional and the social aspects presented in the poll, was measured using the similarity index.

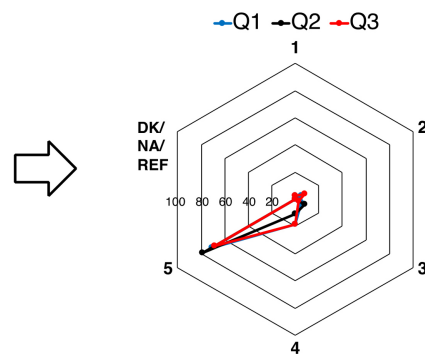
The similarity index was obtained as follows:

1) The percentage of students' responses was plotted as a Cartesian plane plot considering the DK/NA/REF (Don't Know, No Answer, Refusal) alternative of Likert-type scale as a value of 6 (see the left bottom corner of **Figure 1**) for all the formative assessments (see **Figure 2** and **Figure 3**). In the example of **Figure 1**, the top part shows the Likert-type scale responses (left) and the spider chart plot of **Figure 2(c)** (right), for comparison.

2) The Cartesian plane plot of **Figure 1** (right bottom corner) was plot using left bottom corner table of **Figure 1**, as an irregular polygon by closing the shape between the points 1 and 6.

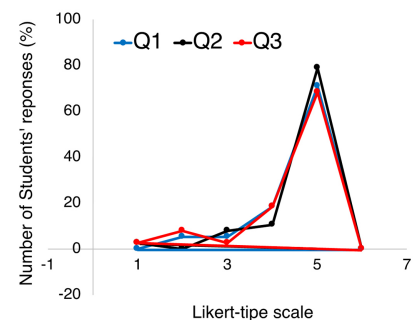
Percentage of students' responses

	Q1	Q2	Q3
1	0	3	3
2	5	0	8
3	5	8	3
4	18	11	18
5	71	79	68
DK/NA/REF	0	0	0

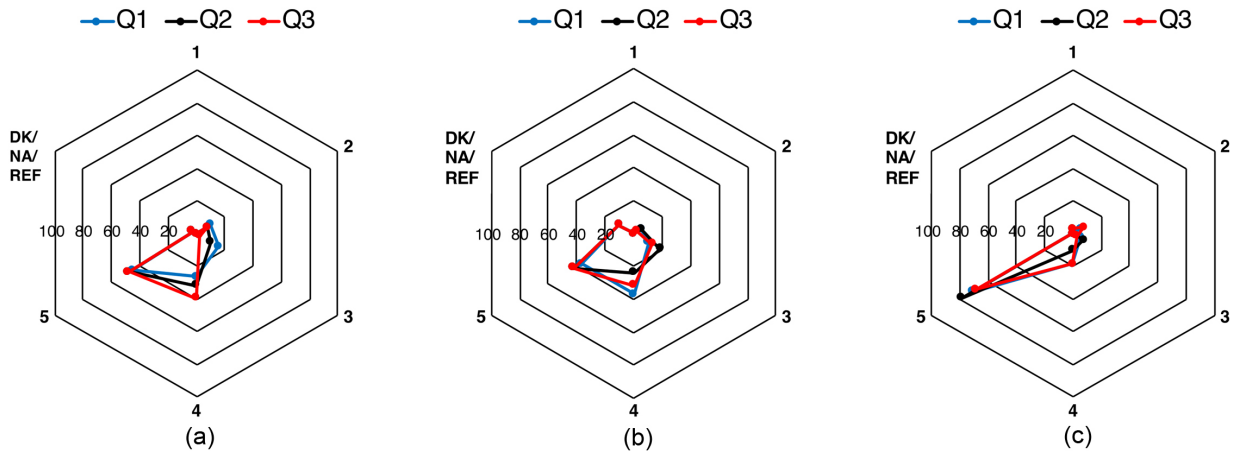


Percentage of students' responses including the Likert-type scale 6 value

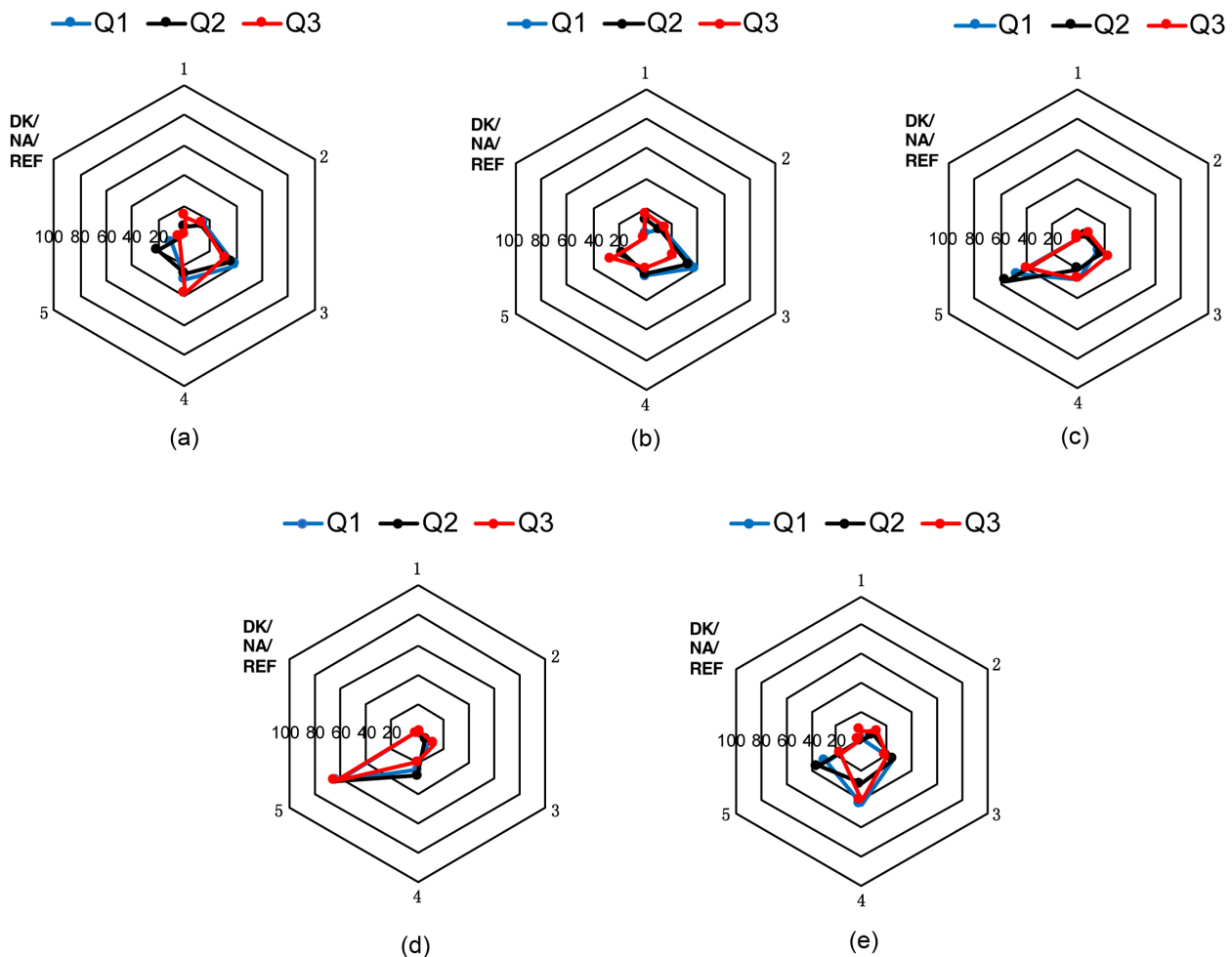
	Q1	Q2	Q3
1	0	3	3
2	5	0	8
3	5	8	3
4	18	11	18
5	71	79	68
6	0	0	0



**Figure 1.** Scheme of transformation from the Likert-type scale responses to the Cartesian plane plot for **Figure 2(c)** (shown in the right upper corner). Q1, Q2, and Q3 stand for question 1 (cognitive aspect, blue line), 2 (social aspect, black line) and 3 (emotional aspect, red line), respectively (see the Poll section).



**Figure 2.** Spider chart plot for the three questions Q1 (cognitive aspect, blue line), Q2 (social aspect, black line), Q3 (emotional aspect, red line) asking in the poll for the following activities: One-minute paper, (a); Directed paraphrasing, (b); General questions/answers out loud, (c) (DK/NA/REF = Don't Know/No Answer/Refusal). All the data are in percentage.



**Figure 3.** Spider chart plot for the three questions Q1 (cognitive aspect, blue line), Q2 (social aspect, black line), Q3 (emotional aspect, red line) asking in the poll for the following activities: One-minute paper, (a); Directed paraphrasing, (b); General questions/answers out loud, (c); POE, (d); Short multiple-choice questions, (e) (DK/NA/REF = Don't Know/No Answer/Refusal). All the data are in percentage.



3) The polygon area was calculated using the method of the Gauss' determinant (Equation (1)) using a counter clockwise order for the three polygons of the Cartesian plane plot (Figure 1, right bottom corner).

$$\text{Area} = \frac{1}{2} \left( \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} + \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} + \dots + \begin{vmatrix} x_N & x_1 \\ y_N & y_1 \end{vmatrix} \right) \quad (1)$$

The colours of the "QX area" (being Q = question, and X = 1, 2 or 3) calculation presented below, match the colours of the Cartesian plane plot of Figure 1:

$$\text{Q1 area} = \frac{1}{2} \left( \begin{vmatrix} 6 & 5 \\ 0 & 71 \end{vmatrix} + \begin{vmatrix} 5 & 4 \\ 71 & 18 \end{vmatrix} + \begin{vmatrix} 4 & 3 \\ 18 & 5 \end{vmatrix} + \begin{vmatrix} 3 & 2 \\ 5 & 5 \end{vmatrix} + \begin{vmatrix} 2 & 1 \\ 5 & 0 \end{vmatrix} + \begin{vmatrix} 1 & 6 \\ 0 & 0 \end{vmatrix} \right) = 99.0 \text{ length}^2$$

$$\text{Q2 area} = \frac{1}{2} \left( \begin{vmatrix} 6 & 5 \\ 0 & 79 \end{vmatrix} + \begin{vmatrix} 5 & 4 \\ 79 & 11 \end{vmatrix} + \begin{vmatrix} 4 & 3 \\ 11 & 8 \end{vmatrix} + \begin{vmatrix} 3 & 2 \\ 8 & 0 \end{vmatrix} + \begin{vmatrix} 2 & 1 \\ 0 & 3 \end{vmatrix} + \begin{vmatrix} 1 & 6 \\ 3 & 0 \end{vmatrix} \right) = 92.0 \text{ length}^2$$

$$\text{Q3 area} = \frac{1}{2} \left( \begin{vmatrix} 6 & 5 \\ 0 & 68 \end{vmatrix} + \begin{vmatrix} 5 & 4 \\ 68 & 18 \end{vmatrix} + \begin{vmatrix} 4 & 3 \\ 18 & 3 \end{vmatrix} + \begin{vmatrix} 3 & 2 \\ 3 & 8 \end{vmatrix} + \begin{vmatrix} 2 & 1 \\ 8 & 3 \end{vmatrix} + \begin{vmatrix} 1 & 6 \\ 3 & 0 \end{vmatrix} \right) = 99.0 \text{ length}^2$$

4) The similarity of the Q1, Q2 and Q3 polygons were obtained by the ratio of their respective areas (Thales' theorem) as is depicted in Table 2. An average of these three measures was calculated. The standard deviation was calculated and expressed as a percentage, and the similarity index was obtained by the subtraction of 100, minus the percentage of standard deviation. This result was called the similarity index.

5) The similarity indexes of the other formative assessments for the first and second term are in the S. I. (see Table C1 and Table C2, Appendix C).

As can be noticed, this algorithm is a simple way to measure the coherence among the mentioned aspects. For that, it is proposed as a direct and a "simple" manner to measure the people's choices regarding polls. For this paper, it was applied to the students' opinion regarding formative assessment.

Combining the results of the Likert-type scale with the similarity index, the situations that could emerge are three:

1) A high Likert-type scale value and a high similarity index value means that the corresponding assessment is well received by the students and should be maintained.

2) A high Likert-type scale value and a low similarity index value (or vice-versa) means that even though the assessment is well rated by the students, there is no coherence among the cognitive, the emotional and the social aspect. As a result, if it is planned to use the instrument again, it should be reformulated. Otherwise, it should be not used again.

**Table 2.** The ratio of the areas of the polygons and the similarity index.

$\frac{\text{Q1 area}}{\text{Q2 area}}$	$\frac{\text{Q1 area}}{\text{Q3 area}}$	$\frac{\text{Q2 area}}{\text{Q3 area}}$	Average	Standard deviation	Percentage of standard deviation (%)	Similarity index (%)
1.08	1.09	1.01	1.06	0.0414	3.91	96.1

3) A low Likert-type scale value, and a low similarity index value means that the assessment should not be used again unless the meaning, the indications and the purpose were presented clearly by the teacher.

### 3. Results and Discussion

As a reminder, the aim of this paper, as is entitled, is collecting the students' preferences through a correlation between three aspects:

- The cognitive aspect → the understanding of the contents → Q1
- The social aspect → the usage of formative evaluations for freshmen → Q2
- The emotional aspect → the students' taste → Q3

The Likert-type scale (Carifio & Perla, 2007; Likert, 1932; Matas, 2018) was used to obtain that information.

This part of the manuscript is devoted to analyse the students' responses to the poll regarding how they evaluated the assessment. The results for the first term, and then, for the second term are presented.

#### 3.1. First Term

**Figure 2** is a spider chart plot for the three questions listed in Methods concerning three formative evaluations: one-minute paper (OMP), general questions/answers out loud (GQAOL) and directed paraphrasing (DP). It is possible to see two things from **Figure 2**. Most of the students oriented their responses to four- and five-values of the Likert-type scale, and there is an overlap between questions Q1, Q2, and Q3 within specific nuances (discussed later). Since five and four received the highest scores, the discussion will be centred on those numbers.

##### One-minute paper

**Figure 2(a)** shows that for the Q1, 45% of the students scored the assessment with five and 26% with four. Adding these percentages, 71% of the students considered this type of assessment equal to or excellent to grasp the course's contents. OMP is an opportunity to review the students' opinion anonymously; they express their beliefs (Such et al., 2015) and they turn so enthusiastic (Hartman & Schachter, 2019). This fact is a fundamental principle that ensures safety in the population when they have polled (Bruschi et al., 2007).

For question Q2 (**Figure 2(a)**, black line), the students declared that 47% would always recommend this evaluation and 32% almost always (79% of the students agree that OMP should be advisable for freshmen).

For question Q3 (**Figure 2(a)**, red line), the sum (4 and 5-values of the Likert-type scale) increases up to 86%. This fact is mostly due to the increase in the percentage of the four-value (38%), comparing to question Q2 (or Q1). Questions as "What did you like most about the lesson?" (the emotional aspect) consider the student's opinions (Such et al., 2015) and their feelings.

Regarding cognitive (Q1) and emotional (Q3) parts, an extract of the student-3's interview could clarify the results:

Teacher: Which one of the following assessments was the one you like the most, questions/answers out loud, one-minute paper or directed paraphrasing?

Student-3: One-minute paper might be... Yes... it might be.

Teacher: Why?

Student-3: Because one could... It's like an instantaneity issue. [...] One could realise immediately what was clear and what wasn't. [...] It's like a... time issue.

Teacher: Okay. So, we could say that you recognise at that moment, your strengths, or weaknesses.

Student-3: Yes.

It seems from the student-3's words that he likes activities in which he obtains rapid satisfaction since that assures him instantly a posture: right or wrong, which allows him to move forward or stop and think.

The students' judgement leads to a similarity index of 87.9% for OMP (**Table 3**) according to the superimposition of Q1, Q2 and Q3 (see **Figure 2(a)**). Even when this value is near to 100%, it is the lowest in comparison to the other ones (see **Table 3**). The cognitive aspect is the responsible for such an "irregularity" (Q1 area = 115, Q2 area = 92.0, Q3 area = 91.0, see **Figure C1**, S. I.). Since the social and the emotional aspect correlated well, it seems the students are not quite sure about the influence of the cognitive aspect. In that regard, OMP should carefully be used, meaning spaced in time and feedbacking the student, class by class.

#### Directed paraphrasing

The results for this issue are depicted in **Figure 2(b)**. Note that the percentages of responses to Q1, Q2, and Q3 are similar among the four and five Likert-type scale values (not the scenario for **Figure 2(a)** results). Likewise, as a general trend, the alternative "sometimes" increases, and the alternative "always" decreases (see **Figure C1**, top, S. I.). It is important to note that another fact that influences this change, is the option DK/NA/REF, which becomes 11% of the answers (see **Figure C1**, top, S. I.).

Even when the sum of four- and five-values are 74%, 66%, and 74% for Q1, Q2, and Q3, respectively (see 4 and 5 entries in **Figure C1**, top, S. I.), the percentage of the five-value decreases regarding those of OMP. It seems the exercise of "being in someone else's shoes" is not wholly accepted by the students. A subject submitted to these kinds of tasks must dominate the concepts, and those concepts, sometimes abstracts, must explain them in ordinary words.

**Table 3.** Similarity index<sup>a</sup> for Q1 (cognitive aspect), Q2 (social aspect), and Q3 (emotional aspect) regarding the assessment of **Figure 2**. All the values are in percentages.

One-minute paper (a)	Directed paraphrasing (b)	General questions/answers out loud (c)
87.9	96.4	96.1

<sup>a</sup>See Methods and **Table C1** in the S. I. for the calculation of this index.

Nevertheless, a coherence of 96.4% (see **Table 2** and **Table C1**, middle, S. I.) among the three aspects was found. It could say that a mixed result arises. Although the similarity index reached the value of 96.4%, the five-value decreases regarding OMP. This counterpoint is noted by the interviews of the student-1 and student-3:

Teacher: Among the three assessment [...] the directed paraphrasing was the lowest scored. Why do you think was like that?

Student 1: We thought it wasn't useful.

Teacher: [...] Could you explain a little bit more?

Student-1: For example, for the other assessments, we could realise what was going to happen. I mean, at the end of the assessment, we had the confirmation or the disconfirmation. But for the paraphrasing... It was like "So what...?"

-----

Teacher: What did you think about this assessment?

Student-3: It was fun...

Teacher: Was it hard, was it easy?

Student-3: It was fun, but it was difficult because one is never in that situation.

[...]

Teacher: [...] I don't know if in the following courses, the teachers use paraphrases, but do you think this activity helped you?

Student-3: Gosh... I think to be more helpful; the assessment should be more repetitive along the semester.

Both students seem shocked by this unusual activity, but student-3 advises about the way to reinforce the usefulness of this activity. Thus, in chemistry subjects, directed paraphrasing must be carefully applied.

#### **General questions/answer out loud**

In **Figure 2(c)**, the results for the GQAOL activity are shown. According to **Table 2**, the overlap between Q1, Q2, and Q3 is 96.1%. Adding five- and four-values for every question, 89% (71% + 18%) of the students agree that GQAOL over the lessons helped them to understand the contents and 90% (79% + 11%) thought should be used for freshmen (see **Figure 2(c)**, blue and black lines, and **Figure C1**, top, S. I.). Regarding taste, 86% (68% + 18%) of the students liked this evaluation very much (see **Figure 2(c)**, red line, and **Figure C1**, top, S. I.). These results demonstrate two things: 1) students recognise GQAOL as necessary, independently of the content. Thus, this assessment should always be used within lessons and 2), considering the students' opinion is a very well valued matter for them (Lancaster, 2007; Wardrop, 2012).

Delving into this point, this is part of the interview for student-1:

Teacher: Why do you think a large part of the class liked this assessment?

Student-1: I think when we did it, we had an epiphany: "Oh, it was for

that!”

Teacher: And at the school, did the teachers ask you “Why do you think this phenomenon is like that?”

Student-1: Nope.

Teacher: They don’t ask you, okay. But did you ask?

Student-1: They say, “Let’s work; let’s work!” “Follow the instructions on the guide.”

Teacher: And for you, is it important that the teacher asks you what do you think?

Student 1: I think it works a lot because immediately, one says: “Oh, now I know it!” And if you don’t, it means something is wrong.

Two things arise from this extract:

1) The students are not used to think from a question (because it is not a regular exercise at the school) and 2) questions are a significant issue for them because, through immediate feedback, it gives them a clue if they are right or wrong (Ruiz-Primo, 2011). This idea agrees with Lancaster (Lancaster, 2007) and with the results for the OMP. Thus, asking students during a lesson is a fast way to move their ideas forward.

### 3.2. Second Term

As was described in the Methods section, during the second term, a group of 23 students were added to the students who had passed the course of GM1. Thus, a total of 84 students enrolled in GM2, but 78 answered the poll.

All the assessment used in the first term were repeated in the second one, and two more were added: POE (Kearney, 2004; Latifah et al., 2019) and short multiple-choice questions (Butler, 2018).

**Figure 3** depicts the spider chart plot for Q1, Q2 and Q3, concerning assessments used in the second term for GM2. At first sight, the correlation between the three questions is not as good for all assessments as in **Figure 2**. A detailed discussion by assessment is presented below.

#### One-minute paper

According to **Figure 3(a)**, for Q1, Q2 and Q3, a similar trend emerges. The four- and five-values of Likert-type scale give a sum of preferences of 39%, 47% and 42%, respectively. In contrast with **Figure 2(a)**, these percentages decrease, showing that a new phenomenon has settled. The incorporation of new students (second time GM2 students + first term students who this time, answered the poll) changed the results. As the survey was anonymous, it is not possible to gather more information.

The similarity index reported in **Table 4** for OMP is one of the lowest for the second term activities. In this case, the slight percentages of the four and five-values correlate well with the low value of the similarity index, indicating that this time, the assessment was not well received. In that regard, it is possible that a part of the first-term students changed their opinion about OMP, and a

**Table 4.** Similarity index<sup>a</sup> for Q1 (cognitive aspect), Q2 (social aspect), and Q3 (emotional aspect) regarding the assessment of **Figure 3**. All the values are in percentages.

One-minute paper (a)	Directed paraphrasing (b)	General questions/answers out loud (c)	POE (d)	Short multiple-choice questions (e)
83.1	81.0	94.2	98.9	91.0

<sup>a</sup>See Methods and **Table C2** in the S. I. for the calculation of this index.

part of the second-term students increased this opinion. This result emphasises feedbacking the students' doubts during the next lesson.

#### Directed paraphrasing

In contrast with the DP of the first term, the DP of the second term was not well judged, 96.4% (see **Table 3**) of similarity, versus 81.0% (see **Table 4**), respectively. The decrease is a consequence of the dispersion among the three aspects caused, principally, by the shape of Q3. Inspecting **Figure 3(b)**, the three-value of Likert-type scale reaches the most significant percentage of preferences for Q1 and Q2 (38% and 33%, respectively). Only for Q3, the five-value (26%) is larger than the others. These results show the students consider that DP does not contribute substantially to the cognitive aspect, nor the social one.

It is necessary to re-evaluate the use of this assessment in future lessons or courses due to, although this instrument is widely used (Bronshiteyn & Baladad, 2006; Cheung, 2016; Moran et al., 2014; Ponce et al., 2012; Tan, 2017; Uemliamin, 2000), a General Chemistry course would not be a proper context for its application (see the first term interviews of the student-1 and the student-3).

#### General questions/answer out loud

In this term, GQAOL were well evaluated by the students as in the first term. According to **Figure 3(c)**, most of the responses were for the four- (27%) and five-values (47%) and, according to the students 74% (27% + 47%) thought that this assessment worked in the cognitive aspect (Q1), 77% (21% + 56%) thought it worked in the social aspect (Q2), and 67% (27% + 40%) believed that it worked for the emotional part (Q3). For GQAOL, the similarity index was 94.2% (see **Table 4**), an indication that the three aspects measured are in consonance. Thus, this assessment should be kept during classes.

Complementing the findings for GQAOL in the first and second term, some thoughts of the student-2 are presented:

Teacher: Why do you think GQAOL helped you?

Student-2: I think one is used to attending lessons, listening to the teacher, taking notes... however, we don't think: "Am I understanding or not?" I think this is the moment [the class] to understand the concepts because, after that, one goes back to home and one has other subjects, and the doubt vanished... And when one is in the exam...

Teacher: "I should have asked this in class..."

Student-2: ¡Right! The moment has already passed.

From the student's thoughts, it is possible to infer that for her, having immediate feedback, again, is extremely important (Havnes et al., 2012). Generalising an idea is the way to know if the students are learning. If not, they pay more attention (or study harder). Alternatively, they fail in the exam.

### **Predict, Observe, Explain**

This assessment has a variety of contexts in which can be applied. In this case, POE was referred to a series of live chemical reactions. As can be seen in **Figure 3(d)**, most of the answers were for the five-value of Likert-type scale. The sum of the percentages of the four-value plus the five-value (see **Figure C2**, top, **Appendix C**, S. I.) gives the following results: 87% (Q1), 92% (Q2) and 83% (Q3). The similarity index reaches a value of 98.9% (see **Table 4**), showing that there is an almost perfect correlation between the cognitive, the social and the emotional part. This result is not surprising, considering that examples of live chemical reactions (the social dimension) are what the students expect to see (the emotional dimension) in a chemistry course. Moreover, chemical reactions are the central part of the course, and for that, the students highly score the cognitive aspect.

Concerning the social aspect, an extract of the student-3's interview adds the following:

Teacher: [...] So, let's move to the last assessment, POE [...]. There was one which took your attention, people, the rainbow solutions...

Student 3: It was fun!

Teacher: Do you think so?

[...]

Student 3: It was a kind of fun because we could chat among the guys in the group, like "What's in there?"

Teacher: Okay.

Student 3: "What's the acid-base indicator?" Or "what acid is in there? What base is in there?" It was like it led to the conversation.

Teacher: Let's say it encouraged the conversation with the others by asking why...

Student 3: Yes.

The student-3 is adding an interesting point of view. The chat among classmates is important because it forces them to discuss. For this chemical reaction, "before the magic happens", the students had to predict what was going to occur, then, observe the colours appearance and finally, write a rational explanation for that phenomenon. The reasonable answer is not a "piece of cake" for the students. They must imagine the sub-microscopical world to explain the macro-phenomenon (change of colour). When a challenge of these characteristics is presented, it puts them in the place of a researcher (Nunez-Oviedo & Clement, 2019). In the end, this is what chemistry is all about, so this might be the reason why POE applied to experiments is highly well evaluated and should be maintained during the course.

### Short multiple-choice questions

SMCQ were used at the end of eight lessons (see **Table B2**, **Appendix B**, S. I.). The results for the three questions appear in **Figure 3(e)**. As it can be seen, the dispersion of the Likert-type scale is more significant in comparison to the other activities. The similarity index reaches a value of 91.0% (see **Table 4**). The sum of the four- and five-values for Q1, Q2, and Q3 is 72%, 64%, and 58%, respectively (see **Figure C2**, top, S. I.). This sum represents the lowest numbers of the four previous assessment and is due to the decrease in the five-value and the increase in the three-value (see **Figure C2**, top, S. I.). In any aspect, SMCQ was the fewer students' favourite. An explanation for this result might be the occasional feedback that was given to the students. The platform Socrative.com was used for this purpose, but since the students not always were able to connect it (for different reasons), they did not receive instant feedback, and they felt disappointed. This fact is in tune with the results for GQAOL.

A correction for this assessment might be the replacement of virtual questions for questions in a piece of paper and the immediate feedback on the whiteboard, for example. Another improvement could be the application of the SMCQ not only at the end but at the beginning of the class, moving the philosophy from a simple test to a pre- and post-test. Pre- and post-test allow comparing knowledge between what the students partially knew or did not know, and what they finally learned.

## 4. Conclusion

A similarity index was created for measuring the coherence between three aspects, cognitive, social, and emotional, in a survey to evaluate assessments. This index gives a direct result of what intends to measure and, it might be extended to evaluate any kind of polls that uses Likert-type scale. Nevertheless, an improvement in which the starting points are those from the spider chart plot, is being carried out for future studies. Moreover, the inclusion of additional questions in the Likert-type polls covering other aspects or applying the index also by student and not only by group is advisable, and we are working on it.

This index evidenced that GQAOL, SMCQ and POE, have the highest coherence according to the students. The students believe that these assessments cognitive, social, and emotional speaking, are necessary. GQAOL is the assessment with the highest similarity index because students obtain instant feedback. Thus, this assessment must be maintained within the lessons.

SMCQ had a fractious reception according to the similarity index, although this fact might be due to a malfunction or the impossibility for accessing to the socrative.com platform. As a result, this assessment should either be eliminated from the lessons or be restructured.

POE was the assessment which obtained the top score in the Likert-type scale and the similarity index. Moreover, POE promotes socio-cognition among students. Thus, this assessment must be maintained within the lessons.



OMP was evaluated relatively well among the students. The reason for that is because this assessment takes into accounts their opinions, but anonymously. Nevertheless, OMP, according to students' opinion, loses its purpose if they do not receive instantaneous feedback.

DP is the lowest-ranked assessment. Besides the small values in the Likert-type scale, the similarity index provides the maximum dispersion. The students do not feel comfortable with this assessment, and they do not understand how should be useful. A re-evaluation of its use must be done for other courses.

In general, the students' opinions led, for this General Chemistry course, to the following increasing ranking of favourite assessments: SMCQ, GQAOL and POE. Reaffirming the words of Chamizo (1995, 1996) and Bulwik (2004) the assessment must be a process more than a goal, where to measure the "students' pulse" along the semester (or along the year) is possible, and not only when the test or the exam arrived. The students appreciate this matter.

Since this is a case of study, it needs validation with a broad quantity of courses, an issue that is in progress.

## Acknowledgements

P.F.M. wants to thank M. Cecilia Nunez-Oviedo for her helpful comments and to the students who participated in this study. They know who they are. Thank you, guys! #RenunciaPiñera #NuevaConstitucion #ChileDespertó #BoricPresidente.

## Supporting Information

Supporting information includes four appendixes in one document:

Appendix A. List of Formative/Summative assessments.

Appendix B. List of content, activity and assessment carried out for every lesson.

Appendix C. Calculation of the similarity index.

Appendix D. References.

## Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

## References

- Angelo, T., & Cross, P. (1993). *Classroom Assessment Techniques*. A Handbook for College Teachers. Jossey-Bass.
- Arriaga, M. O., Villegas, G., & Arciniega, L. (2017). Evaluación del aprendizaje significativo con la aplicación Socrative. Evaluation of Meaningful Learning with the Socrative Application. *Tecnología Educativa Revista Conaic*, 4, 52-57.  
<https://doi.org/10.32671/terc.v4i2.106>
- Bannert, M., & Mengelkamp, C. (2008). Assessment of Metacognitive Skills by Means of Instruction to Think Aloud and Reflect When Prompted. Does the Verbalisation Me-

- thod Affect Learning? *Metacognition and Learning*, 3, 39-58.  
<https://doi.org/10.1007/s11409-007-9009-6>
- Blackburn, M. (2015). I Am Not a Superhero But I Do Have Secret Weapons! Using Technology in Higher Education Teaching to Redress the Power Balance. *Journal of Pedagogic Development*, 5, pages. <http://hdl.handle.net/10547/346546>
- Bresnock, A. E., Graves, P. E., & White, N. (1989). Multiple-Choice Testing—Question and Response Position. *Journal of Economic Education*, 20, 239-245.  
<https://doi.org/10.1080/00220485.1989.10844626>
- Bronshteyn, K., & Baladad, R. (2006). Perspectives on ... Librarians as Writing Instructors: Using Paraphrasing Exercises to Teach Beginning Information Literacy Students. *Journal of Academic Librarianship*, 32, 533-536.  
<https://doi.org/10.1016/j.acalib.2006.05.010>
- Bruschi, D., Lanzi, A., & Fovino, I. N. (2007). A Protocol for Anonymous and Accurate E-Polling. In *Secure E-Government Web Services* (pp. 180-198). IGI Global.  
<https://doi.org/10.4018/978-1-59904-138-4.ch011>
- Bulwik, M. (2004). La evaluación de los aprendizajes y el portafolios. *Educación Química*, 15, 104-107. <https://doi.org/10.22201/fq.18708404e.2004.2.66195>
- Butler, A. C. (2018). Multiple-Choice Testing in Education: Are the Best Practices for Assessment Also Good for Learning? *Journal of Applied Research in Memory and Cognition*, 7, 323-331. <https://doi.org/10.1016/j.jarmac.2018.07.002>
- Carifio, J., & Perla, R. J. (2007). Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and Their Antidotes. *Journal of Social Sciences*, 3, 106-116.  
<https://doi.org/10.3844/jssp.2007.106.116>
- Chamizo, J. A. (1995). Mapas conceptuales en la enseñanza y la evaluación de la química. *Educación Química*, 6, 118-124. <https://doi.org/10.22201/fq.18708404e.1995.2.66719>
- Chamizo, J. A. (1996). Evaluación de los aprendizajes en química. Segunda parte: Registros de aprendizaje, asociación de palabras y portafolios. *Educación Química*, 7, 86-89. <https://doi.org/10.22201/fq.18708404e.1996.2.66671>
- Cheung, A. K. F. (2016). Paraphrasing Exercises and Training for Chinese to English Consecutive Interpreting. FORUMFORUM. *Revue Internationale d'interprétation et de Traduction/International Journal of Interpretation and Translation*, 14, 1-18.  
<https://doi.org/10.1075/forum.14.1.01che>
- Crete, P. (2005). Should Student Learning Journals Be Assessed? *Assessment & Evaluation in Higher Education*, 30, 287-296.  
<https://doi.org/10.1080/02602930500063850>
- Davari, M., & Bahraman, M. (2019). I Explain, Therefore I Learn: Improving Students' Assessment Literacy and Deep Learning by Teaching. *Studies in Educational Evaluation*, 61, 66-73. <https://doi.org/10.1016/j.stueduc.2019.03.002>
- Dodd, D. K., & Leal, L. (1988). Answer Justification—Removing the Trick from Multiple-Choice Questions. *Teaching of Psychology*, 15, 37-38.  
[https://doi.org/10.1207/s15328023top1501\\_8](https://doi.org/10.1207/s15328023top1501_8)
- Fischman, G. E., Topper, A. M., Silova, I., Goebel, J., & Holloway, J. L. (2019). Examining the Influence of International Large-Scale Assessments on National Education Policies. *Journal of Education Policy*, 34, 470-499.  
<https://doi.org/10.1080/02680939.2018.1460493>
- Frías, M. A., Arce, C., & Flores-Morales, P. (2016). Uso de la plataforma socrative.com para alumnos de Química General. *Educación Química*, 27, 59-66.  
<https://doi.org/10.1016/j.eq.2015.09.003>

- Guarascio, A. J., Nemecek, B. D., & Zimmerman, D. E. (2017). Evaluation of Students' Perceptions of the Socrative Application versus a Traditional Student Response System and Its Impact on Classroom Engagement. *Currents in Pharmacy Teaching and Learning*, 9, 808-812. <https://doi.org/10.1016/j.cptl.2017.05.011>
- Güven, E. (2014). The Effect of Project Based Learning Method Supported by Prediction-Observation-Explanations on the Attitude and Behaviors towards Environmental Problems. *Education and Science*, 39, 25-39.
- Hartman, T., & Schachter, E. (2019). High School Students "Talk Back" to the Anonymous Researcher behind the Questionnaire. *The Journal of Experimental Education*, Volume number, 1-13. <https://doi.org/10.1080/00220973.2019.1573795>
- Harwood, W. S. (1996). The One-Minute Paper: A Communication Tool for Large Lecture Classes. *Journal of Chemical Education*, 73, 229. <https://doi.org/10.1021/ed073p229>
- Havnes, A., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative Assessment and Feedback: Making Learning Visible. *Studies in Educational Evaluation*, 38, 21-27. <https://doi.org/10.1016/j.stueduc.2012.04.001>
- Hong, J. C., Hwang, M. Y., Liu, M. C., Ho, H. Y., & Chen, Y. L. (2014). Using a "Prediction-Observation-Explanation" Inquiry Model to Enhance Student Interest and Intention to Continue Science Learning Predicted by Their Internet Cognitive Failure. *Computers and Education*, 72, 110-120. <https://doi.org/10.1016/j.compedu.2013.10.004>
- Hortigüela, D., Palacios, A., & López, V. (2019). The Impact of Formative and Shared or Co-Assessment on the Acquisition of Transversal Competences in Higher Education. *Assessment and Evaluation in Higher Education*, 44, 933-945. <https://doi.org/10.1080/02602938.2018.1530341>
- Hsu, C. Y., Tsai, C. C., & Liang, J. C. (2011). Facilitating Preschoolers' Scientific Knowledge Construction via Computer Games Regarding Light and Shadow: The Effect of the Prediction-Observation-Explanation (POE) Strategy. *Journal of Science Education and Technology*, 20, 482-493. <https://doi.org/10.1007/s10956-011-9298-z>
- Kearney, M. (2004). Classroom Use of Multimedia-Supported Predict-Observe-Explain Tasks in a Social Constructivist Learning Environment. *Research in Science Education*, 34, 427-453. <https://doi.org/10.1007/s11165-004-8795-y>
- Kokina, J., & Juras, P. E. (2017). Using Socrative to Enhance Instruction in an Accounting Classroom. *Journal of Emerging Technologies in Accounting*, 14, 85-97. <https://doi.org/10.2308/jeta-51700>
- Kose, S., & Bilen, K. (2012). The Effect of Laboratory Activities Based on a POE Strategy for Class Teacher Candidates' Achievements, Science Process Skills and Understanding the Nature of Science. *Energy Education Science and Technology Part B Social and Educational Studies*, 4, 2357-2368.
- Kumar, A., Kumar, M. S., & Gulia, R. (2017). The Use of Classroom Assessment Techniques (CATs) to Enhance Learning and Motivation in First Year Undergraduate Students. *Journal of Research in Medical Education & Ethics*, 7, 93. <https://doi.org/10.5958/2231-6728.2017.00016.6>
- Lancaster, E. L. (2007). Questions and Answers (Encouraging Students to Sing and Count out Loud). *Clavier*, 46, 48.
- Latifah, S., Irwandani, I., Saregar, I., Diani, R., Fiani, O., Widayanti, W., & Deta, U. A. (2019). How the Predict-Observe-Explain (POE) Learning Strategy Remediate Students' Misconception on Temperature and Heat Materials? *Journal of Physics: Conference Series*, 1171, 1-6. <https://doi.org/10.1088/1742-6596/1171/1/012051>
- Lewis, C. E. (2019). Is the Flipped Classroom a Panacea for Medical Education? *Current*

- Surgery Reports*, 7, 9. <https://doi.org/10.1007/s40137-019-0230-4>
- Likert, R. (1932). A Technique for the Measurement of Attitude. *Archives of Psychology*, 140, 5-55.
- Lutterodt, M. C. (2017). Qualitative Formative Feedback to the Teacher. The Use of “Learn Evaluation” a Modified “One Minute Paper”. In *Improving University Science Teaching and Learning—Pedagogical Projects 2017* (pp. 133-147). Department of Science Education, University of Copenhagen.
- Manning, R. D., Keiper, M. C., & Jenny, S. E. (2017). Pedagogical Innovations for the Millennial Sport Management Student: Socratic and Twitter. *Sport Management Education Journal*, 11, 45-54. <https://doi.org/10.1123/smej.2016-0014>
- Matas, A. (2018). Diseño del formato de escalas tipo Likert: Un estado de la cuestión. *Revista Electronica de Investigacion Educativa*, 20, 38-47. <https://doi.org/10.24320/redie.2018.20.1.1347>
- Maxwell, S. R. J. (2012). An Agenda for UK Clinical Pharmacology. How Should Teaching of Undergraduates in Clinical Pharmacology and Therapeutics Be Delivered and Assessed? *British Journal of Clinical*, 73, 893-899. <https://doi.org/10.1111/j.1365-2125.2012.04232.x>
- Mayowski, C. A., Norman, M. K., & Kapoor, W. N. (2018). Assessing an Assessment: The Review and Redesign of a Competency-Based Mid-Degree Evaluation. *Journal of Clinical and Translational Science*, 2, 223-227. <https://doi.org/10.1017/cts.2018.321>
- Metcalfe, J., & Xu, J. (2018). Learning from One’s Own Errors and Those of Others. *Bulletin & Review*, 25, 402-408. <https://doi.org/10.3758/s13423-017-1287-7>
- Moran, A. S., Swanson, H. L., Gerber, M. M., & Fung, W. (2014). The Effects of Paraphrasing Interventions on Problem-Solving Accuracy for Children at Risk for Math Disabilities. *Learning Disabilities Research & Practice*, 29, 97-105. <https://doi.org/10.1111/ldrp.12035>
- Nunez-Oviedo, M. C., & Clement, J. J. (2019). Large Scale Scientific Modeling Practices That Can Organize Science Instruction at the Unit and Lesson Levels. *Frontiers in Education*, 4, 1-22. <https://doi.org/10.3389/educ.2019.00068>
- Orozco, H. O., Sosa, M. R., & Martínez, F. (2018). Didactic Models at the High Education: A Reality That Can Change. *Profesorado—Revista de Curriculum y Formación Del Profesorado*, 22, 405-427.
- Ponce, H. R., López, M. J., & Mayer, R. E. (2012). Instructional Effectiveness of a Computer-Supported Program for Teaching Reading Comprehension Strategies. *Computers and Education*, 59, 1170-1183. <https://doi.org/10.1016/j.compedu.2012.05.013>
- Prashanti, E., & Ramnarayan, K. (2019). Ten Maxims of Formative Assessment. *Advances in Physiology Education*, 43, 99-102. <https://doi.org/10.1152/advan.00173.2018>
- Price, N., Stephens, A. L., Clement, J., & Nunez-Oviedo, M. C. (2017). Using Imagery Support Strategies to Develop Powerful Imagistic Models. *Science Scope*, 41, 40-49. [https://doi.org/10.2505/4/ss17\\_041\\_04\\_40](https://doi.org/10.2505/4/ss17_041_04_40)
- Quesada, V., Gómez, M. Á., Gallego, M. B., & Cubero-Ibáñez, J. (2019). Should I Use Co-Assessment in Higher Education? Pros and Cons from Teachers and Students’ Perspectives. *Assessment & Evaluation in Higher Education*, 44, 987-1002. <https://doi.org/10.1080/02602938.2018.1531970>
- Robertson, J. (2012). Likert-Type Scales, Statistical Methods, and Effect Sizes. *Communications of the ACM*, 55, 6. <https://doi.org/10.1145/2160718.2160721>
- Rubenson, K. (2019). Assessing the Status of Lifelong Learning: Issues with Composite

- Indexes and Surveys on Participation. *International Review of Education*, 65, 295-317. <https://doi.org/10.1007/s11159-019-09768-3>
- Ruiz-Primo, M. A. (2011). Informal Formative Assessment: The Role of Instructional Dialogues in Assessing Students' Learning. *Studies in Educational Evaluation*, 37, 15-24. <https://doi.org/10.1016/j.stueduc.2011.04.003>
- Schneider, M. C., McDonel, J. S., & DePascale, C. A. (2019). Performance Assessment and Rubric Design. In *The Oxford Handbook of Assessment Policy and Practice in Music Education* (Volume 1, pages). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190248093.013.27>
- Smith, S. (2007). How Does Student Performance on Formative Assessments Relate to Learning Assessed by Exams? *Journal of College Science Teaching*, 36, 28-34.
- Stead, D. R. (2005). A Review of the One-Minute Paper. *Active Learning in Higher Education*, 6, 118-131. <https://doi.org/10.1177/1469787405054237>
- Stewart, S., & Richardson, B. (2000). Reflection and Its Place in the Curriculum on an Undergraduate Course: Should It Be Assessed? *Assessment & Evaluation in Higher Education*, 25, 369-380. <https://doi.org/10.1080/713611443>
- Such, J. M., Criado, N., & García-Fornes, A. (2015). An Active Learning Technique Enhanced with Electronic Polls. *International Journal of Engineering Education*, 31, 1048-1057.
- Tan, C. (2019). PISA and Education Reform in Shanghai. *Critical Studies in Education*, 60, 391-406. <https://doi.org/10.1080/17508487.2017.1285336>
- Tan, K. E. (2017). Using Online Discussion Forums to Support Learning of Paraphrasing. *British Journal of Educational Technology*, 48, 1239-1249. <https://doi.org/10.1111/bjet.12491>
- Uemlianin, I. A. (2000). Engaging Text: Assessing Paraphrase and Understanding. *Studies in Higher Education*, 25, 346-358. <https://doi.org/10.1080/713696160>
- Viera, L., Ramírez, S., Wainmier, C., & Salinas, J. (2007). Criterios y actividades para la evaluación del aprendizaje en cursos universitarios de química. *Educación Química*, 18, 294-302. <https://doi.org/10.22201/fq.18708404e.2007.4.65876>
- Wardrop, J. (2012). Speaking out Loud: Muslim Women, Indian Delights and Culinary Practices in eThekweni/Durban. *Social Dynamics—A Journal of African Studies*, 38, 221-236. <https://doi.org/10.1080/02533952.2012.717209>
- White, R., & Gunstone, R. (1992). *Probing Understanding*. The Falmer Press.

## Appendix A: List of Formative/Summative Assessments

- Chain notes (Keeley, 2015)
- Checklist (Baldwin & Ching, 2019; Caruso et al., 2019; Dreimuller et al., 2019; Robles, 2019)
- Directed paraphrasing (Angelo & Cross, 1993; Keeley, 2015)
- Double-entry journal (Allen, 2008; Angelo & Cross, 1993; Berthoff, 1982)
- Group and Self-assessment (Bartels & Kulgemeyer, 2019; Herrington & Sweeder, 2018; Sridharan & Boud, 2019; To & Panadero, 2019)
- In basket (del Pozo, 2012; Frederiksen et al., 1957; Schippmann et al., 1990)
- KPSI (Knowledge and Prior Study Inventory) (Quintanilla et al., 2008; Tamir & Lunetta, 1978)
- K-W-L (Know-Want-Learned) chart (Ogle, 1986)
- Learning registers (Chamizo, 1996)
- Mental/conceptual maps (Chamizo, 1995; Pendley et al., 1994)
- One-minute paper (Angelo & Cross, 1993)
- One-sentence summary (Berthoff, 1982)
- One-world journal (Angelo, 1991)
- POE (Prediction-Observation-Explanation) (Jasdilla et al., 2019; Kearney, 2004; Latifah et al., 2019; Treagust & Chong, 1995)
- Portfolios (Fosado et al., 2018; Lyons, 1999)
- Question/answer out loud (Bannert & Mengelkamp, 2008; Lancaster, 2007; Metcalfe & Xu, 2018; Wardrop, 2012)
- RSQC2 (Recall, Summarize, Question, Connect, Comment) (Angelo & Cross, 1993; Cowan & George, 2004)
- Rubrics (Angra & Gardner, 2018; Cheng & Chan, 2019; Cockett & Jackson, 2018; Tobajas et al., 2019)
- (Short) multiple-choice questions (Bresnock et al., 1989; Butler, 2018; da Silva, 2019; Dodd & Leal, 1988; Witchel et al., 2018)
- Venn diagrams (Hatzikiriakou & Metallidou, 2009; Kerr & Macosko, 2011; Wygoda & Tague, 1995)
- Words association (Sutton, 1980; Zakaluk et al., 1986)

## Appendix B

**Table B1.** List of content, activity and assessment carried out for every lesson during general chemistry 1.

Class	Content <sup>a</sup>	Activity <sup>b</sup>	Assessment
1	Presentation of the course	-	-
2	Fundamental concepts about matter and measurements. Classification, chemical and physical properties, and states of matter	Students, in groups, analyse substances of everyday life, smelling, tasting, and touching them	Groups discuss and point out what type of substance they have? An element, a molecule, an ionic compound, or a mixture?
3	Quantitative measurements and the metric system of unities. Scientific notation, significant figures. Dimensional analysis	Experiment: "Plastic cube with sand and water"	<i>General question/answer out loud</i>
4	Structure of matter and inorganic nomenclature. Early atomic models, atomic theory, and atom structure. Atomic number, atomic weight, isotopes. Periodic Table	Exposition	<i>One-minute paper</i>
5	Seminar 1. Lessons 2 to 4.	Exercises	
6	Ions, molecules. Empiric and molecular formulae. Nomenclature of ionic compounds, molecules, acids and bases	Exposition	<i>One-minute paper</i>
7	Atomic structure. The electromagnetic spectrum, black body radiation, photoelectrical effect, Planck's quantum theory	Video: "Atomic models"	<i>Short multiple-choice questions in socrative.com</i>
8	Seminar 2. Lessons 6 and 7.	Exercises	
9	Line spectra of atoms. Rydberg's equation. Bohr's atomic model and its limitations	Exposition	<i>Short multiple-choice questions in socrative.com</i>
10	Atomic quantum model: de Broglie's postulate, Heisenberg's uncertainty principle and Schrödinger's equation. Orbitals and quantic numbers	Analogy: "A hummingbird in the forest" for the atomic quantum model	<i>Short questions</i>
11	Test 1. Lessons 2 to 8		
12	Atomic structure. Electronic configurations of metals and ions. Valence electrons. Diamagnetism and paramagnetism	Exposition	<i>Short multiple-choice questions in socrative.com</i>
13	Periodic trends. Periods, groups, and Periodic Law (periodic table). Metals, non-metals, semi-metals. Representative elements, noble gases, and transition metals	Exposition	-
14	Seminar 3. Lessons 9 to 12	Exercises	
15	Periodic table. Effective nuclear charge, atomic radii, ionic radii, ionization energy, electron affinity	Exposition	<i>Short multiple-choice questions in socrative.com</i>

## Continued

16	Chemical properties of some chemical groups. Chemical bond. Lewis symbols, octet rule. Ionic bond, covalent bond, resonance hybrids	Exposition	-
17	Seminar 4. Lessons 15 and 16	Exercises	
18	Dipole moments. Prediction of molecular geometries. Hybridisation. Molecular orbitals	Application: "Molecular shape" from Phet <sup>c</sup>	Students assign the molecular geometry to a set of molecules
19	Intermolecular forces. Dipole-dipole attraction, ion-dipole attraction, van der Waal forces, London forces, hydrogen bond	Searching on the internet the principal intermolecular forces	<i>Short multiple-choice questions in socrative.com</i>
20	Seminar 5. Lessons 18 and 19	Exercises	
21	Gases. Properties of the gases. Boyle's and Charles's Law, Avogadro's law. Ideal-Gas equation	Students interact with a balloon	Students answer a multiple-choice questionnaire through the <i>socrative.com</i>
22	Dalton's law. Kinetic theory, Graham's law of effusion, van der Waals equation	Exposition	-
23	Test 4. Lessons 12 to 15 and 17 to 20		
24	Properties of liquids. Structure and properties of water. Solids, crystal structures	Exposition	<i>One-minute paper</i>
25	X-ray diffraction. Heating curves, phase diagrams	Exposition	-
26	Seminar 6. Lessons 21, 22, 24 and 25	Exercises	
27	Stoichiometry. The mol. Determination of the empirical and molecular formula	Using coins to represent atoms and molecules	<i>Short multiple-choice questions in socrative.com</i>
28	Reactions and chemical equations. Types of chemical reactions	Exposition	-
29	Seminar 7. Lessons 27 y 28	Exercises	
30	Stoichiometric calculations,	Video: Chemical reaction	<i>One-minute paper</i>
31	Limiting reactant, percentage yield	Exposition	-
32	Seminar 8. Lessons 30 and 31	Exercises	
33	Solutions. Definition, solubility, concentration expressions	Exposition	<i>One-minute paper</i>
34	Exercises based on concentration expressions	Exposition	-
35	Test 5. Lessons 24 to 32		
36	Types of solutes (non-electrolyte, strong and weak electrolytes). Colligative properties	Experiment: "Electrical conductance"	<i>Directed paraphrasing</i>
37	Henry's law. Exercises based on colligative properties	Exposition	-
	Seminar 9. Lessons 37, 39 and 40	Exercises	



## Continued

39	Reactions in aqueous solution. Acid-base reactions; neutralisation and titration	Experiment: "Acid-base and precipitation reactions"	<i>POE assessment</i>
40	Oxidation-reduction reactions, Oxidation-reduction titrations	Experiment: "oxidation-reduction reaction"	<i>One-minute paper</i>
41	Seminar 10. Lessons 37, 39 and 40	Exercises	
42	Thermochemistry. Basic definitions. Energy, work and heat. Law zero	Demonstration: an experiment of an exothermic reaction	<i>One-minute paper</i>
43	First law. Enthalpy, enthalpy changes	Exposition	<i>One-minute paper</i>
44	Seminar 11. Lessons 41 and 42	Exercises	
45-46	Thermochemistry. Calorimetry. Enthalpy of formation, enthalpy of reaction, the enthalpy change of combustion and enthalpy change of solution. Hess's Law	Exposition	<i>One-minute paper</i>
47	Test 6. Lessons 33 to 42		

<sup>a</sup>Based on the course syllabus. <sup>b</sup>All the lessons included a lecture. Thus, when no activity has been indicated, the evaluation was based on the lecture. Phet, interactive simulations from the University of Colorado, <https://phet.colorado.edu/es/simulations/category/chemistry>.

**Table B2.** List of content, activity and assessment carried out for every lesson during general chemistry 2.

Class	Content <sup>a</sup>	Activity <sup>b</sup>	Assessment
1	Presentation of the course	-	-
2	Chemical thermodynamics. Spontaneous processes. Entropy and the second law of thermodynamics. Gibbs' free energy. Free energy and chemical equilibrium	Exposition	-
3	Chemical kinetics. Rate of a reaction and stoichiometry. Reaction rate law, reaction order and its experimental obtention. The relation between reaction concentration and time: first-order reactions and its half-lives	The teacher introduces the concept of the reaction rate with a pile of sheets and a stapler	<i>General question/answer out loud</i>
4	First-order reactions and its half-lives. Second-order reactions and its half-lives. Effects that influence a reaction-rate. Collision theory and transition state. Effect of the concentration in collisions.	Experiment: "Effervescent tablet in the water"	<i>One-minute paper</i>
5	Activation energy. Arrhenius' equation. Chemical mechanisms and molecularity. Catalysis	Experiment: decomposition of hydrogen peroxide	<i>One-minute paper</i>
6	Seminar 1. Lessons 2 to 5	Exercises	Students answer the exercises
7	Chemical Equilibria. The concept of equilibrium. The chemical equilibrium constant ( $K_C$ ). The relation between $K_C$ and $K_p$ . Multiple equilibriums	Analogy: exchange of students from one side of the classroom to the other one.	Students answer a poll

## Continued

8	Test 1. Lessons 2 to 6.		
9	Homogeneous and heterogeneous equilibriums. Calculations of the concentrations at equilibrium	Exposition	<i>General question/answer out loud</i>
10	The factors that affect chemical equilibrium. Le Chatelier's principle. The response of equilibria to changes in conditions	Experiment: traffic light of solutions	Students answer a poll
11	Acid-Base Equilibria. Definition of acid and base. Arrhenius acids and bases. Brønsted-Lowry acids and bases. Lewis acids and bases. The acid-base properties of water. The pH scale	Experiment: "Rainbow solutions"	<i>POE assessment</i>
12	Strength of acid and bases. Solutions of strong acids and bases. Solutions of weak acids and bases.	Exposition	<i>One-minute paper</i>
13	Buffer solutions	Exposition	<i>One-minute paper</i>
14	Seminar 3. Lessons 7, 9 to 12	Exercises	Students answer the exercises
15	Salts and hydrolysis	Exposition	<i>Short multiple-choice questions in socrative.com</i>
16	Test 2. Lessons 7, 9 to 14		
17	Acid-base indicators. Strong acid-strong base titrations. Strong acid-weak base and weak acid-strong base titrations	Exposition	<i>Short multiple-choice questions in socrative.com</i>
18	Solubility Equilibria. Importance. Solubility and molar solubility. The solubility product ( $K_p$ )	Mental experiment: Solubility equilibria in the eggshells	<i>Directed paraphrasis</i>
19	The common-ion effect. Predicting precipitation. Selective precipitation	Exposition	<i>Short multiple-choice questions in socrative.com</i>
20	Seminar 3. Lessons 15, 17 to 19	Exercises	
21	Sulphurs precipitation. Factors that affect solubility equilibria	-	-
22	Equilibrium of complex ion formation. History overview. Chemical structure of the metal complex. Basic nomenclature of the complex ions. Complex ion equilibria	Experiment: Silver complex compound formation	<i>One-minute paper</i>
23	Seminar 4. Lessons 21 and 22	Exercises	
24	Electrochemistry. Definition. Representation of redox reactions. Galvanic cells. Standard potentials	Link for the previous study: <a href="http://cienciasofa.com/2013/07/respuestas-xix-anguilas-electricas.html">http://cienciasofa.com/2013/07/respuestas-xix-anguilas-electricas.html</a>	<i>Short multiple-choice questions in socrative.com</i>
25	Test 3. Lessons 15, 17 to 23		
26	Cell potential and reaction Gibbs free energy	-	-

## Continued

27	The Nernst Equation. Concentration cells. Batteries	Exposition	<i>One-minute paper</i>
28	Seminar 5. Lessons 26 and 27	Exercises	
29	Corrosion. Electrolytic cells	Experiment: Water electrolysis	<i>Short multiple-choice questions in socrative.com</i>
30	Test 4. Lessons 24, 26 to 28.		
31	Applications of Electrolysis (metallurgic of some metals)	Experiment: Galvanized of a spoon	<i>One-minute paper</i>
32	Organic compounds. General characteristics of organic molecules. Hydrocarbons: alkanes	Videos: watching three ads involving organic substances	<i>Short multiple-choice questions in socrative.com</i>
33	Seminar 6. Lessons 29, 31 and 32	Exercises	
34	Cycloalkanes. Reactions of alkanes. Alkenes and alkynes. Aromatic compounds	-	-
35	Reactions of electrophilic addition, electrophilic substitution, and electrophilic elimination	Exposition	<i>One-minute paper</i>
36	Alcohols, ethers, aldehydes, and ketones	-	-
37	Carboxylic acids, esters, amines, and amides	Experiment: "stink + travertine = perfume"	<i>Short multiple-choice questions in socrative.com</i>
38	Seminar 7. Lessons 34 to 37	Exercises	
39	Test 5. Lessons 31 to 36		
40	Aromatic compounds, quirkality	-	-
41	Synthetic and natural polymers	Conference of a faculty member	<i>General question/answer out loud</i>
42	Nucleic acids	Conference of a faculty member	<i>General question/answer out loud</i>
43	Amino acids and proteins	Conference of a faculty member	<i>General question/answer out loud</i>
44	Seminar 8. Lessons 45 and 46	Exercises	
45	Nuclear chemistry. Radioactivity	Links for the previous study: <a href="http://www.dailymotion.com/video/x6t846">http://www.dailymotion.com/video/x6t846</a> <a href="https://vimeo.com/75872397">https://vimeo.com/75872397</a> <a href="http://www.rtve.es/alacarta/videos/centro-medico/centro-medico-09-11-15/3354227/">http://www.rtve.es/alacarta/videos/centro-medico/centro-medico-09-11-15/3354227/</a>	<i>Short multiple-choice questions in socrative.com</i>

**Continued**

---

46	Nuclear decay	-	-
47	Seminar 10. Lessons 40 to 43		Exercises
48	Test 6. Lessons 40 to 46		
49	Fission and fusion. The impact on biology	-	-

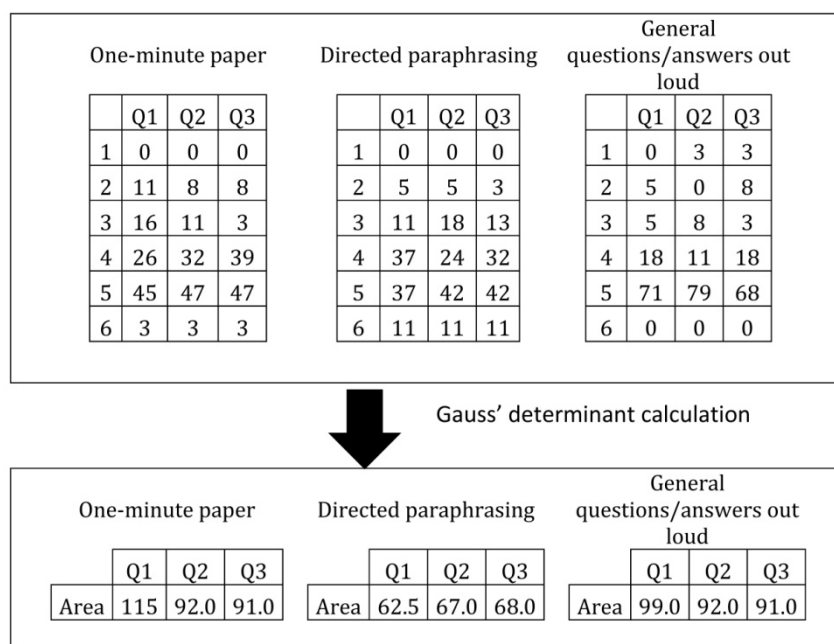
---

<sup>a</sup>Based on the course syllabus. <sup>b</sup>All the lessons included a lecture. Thus, when no activity has been indicated, the evaluation was based on the lecture.

## Appendix C: Similarity Index

The similarity index was calculated, as was described in Methods. Here, a summary of the calculations for all the formative evaluations (first and second term) is presented.

### First-term

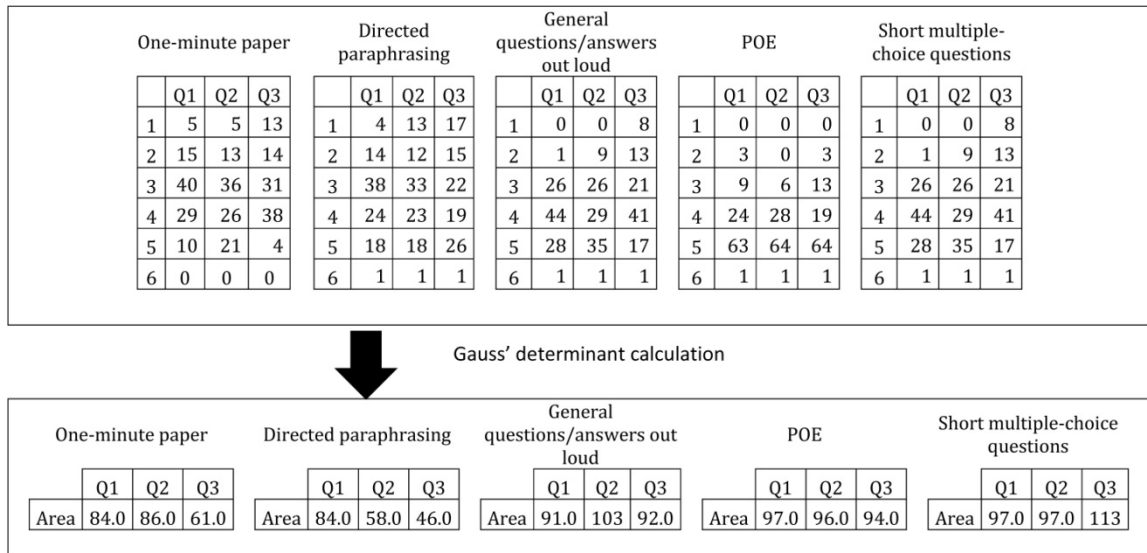


**Figure C1.** Scheme of transformation from modified Likert-type scale responses to polygons' areas for the first-term assessments. All the values are in percentage.

**Table C1.** Areas' ratio of the polygons and the similarity index for the first term: One minute-paper, (a); Directed paraphrasing, (b); General questions/answers out loud, (c).

(a)						
$\frac{Q1 \text{ area}}{Q2 \text{ area}}$	$\frac{Q1 \text{ area}}{Q3 \text{ area}}$	$\frac{Q2 \text{ area}}{Q3 \text{ area}}$	Average	Standard deviation	Percentage of standard deviation	Similarity index (%)
1.25	1.26	1.01	1.17	0.142	12.1	87.9
(b)						
$\frac{Q2 \text{ area}}{Q1 \text{ area}}$	$\frac{Q3 \text{ area}}{Q1 \text{ area}}$	$\frac{Q3 \text{ area}}{Q2 \text{ area}}$	Average	Standard deviation	Percentage of standard deviation	Similarity index (%)
1.07	1.09	1.01	1.06	0.0384	3.63	96.4
(c)						
$\frac{Q1 \text{ area}}{Q2 \text{ area}}$	$\frac{Q1 \text{ area}}{Q3 \text{ area}}$	$\frac{Q2 \text{ area}}{Q3 \text{ area}}$	Average	Standard deviation	Percentage of standard deviation	Similarity index (%)
1.08	1.09	1.01	1.06	0.0414	3.91	96.1

Second-term



**Figure C2.** Scheme of transformation from modified Likert-type scale responses to polygons' areas for the second term assessments. All the values are in percentage.

**Table C2.** Areas' ratio of the polygons and the similarity index for the second term: One minute-paper, (a); Directed paraphrasing, (b); General questions/answers out loud, (c); POE, (d); Short multiple-choice questions, (e).

(a)							
$\frac{Q2 \text{ area}}{Q1 \text{ area}}$	$\frac{Q1 \text{ area}}{Q3 \text{ area}}$	$\frac{Q2 \text{ area}}{Q3 \text{ area}}$	Average	Standard deviation	Percentage of standard deviation	Similarity index (%)	
1.02	1.38	1.41	1.27	0.214	16.9	83.1	
(b)							
$\frac{Q1 \text{ area}}{Q2 \text{ area}}$	$\frac{Q1 \text{ area}}{Q3 \text{ area}}$	$\frac{Q2 \text{ area}}{Q3 \text{ area}}$	Average	Standard deviation	Percentage of standard deviation	Similarity index (%)	
1.45	1.83	1.26	1.51	0.288	19.0	81.0	
(c)							
$\frac{Q2 \text{ area}}{Q1 \text{ area}}$	$\frac{Q3 \text{ area}}{Q1 \text{ area}}$	$\frac{Q2 \text{ area}}{Q3 \text{ area}}$	Average	Standard deviation	Percentage of standard deviation	Similarity index (%)	
1.13	1.01	1.11	1.08	0.0634	5.85	94.2	
(d)							
$\frac{Q1 \text{ area}}{Q2 \text{ area}}$	$\frac{Q1 \text{ area}}{Q3 \text{ area}}$	$\frac{Q2 \text{ area}}{Q3 \text{ area}}$	Average	Standard deviation	Percentage of standard deviation	Similarity index (%)	
1.01	1.03	1.02	1.02	0.0107	1.05	98.9	

(e)

$\frac{Q1 \text{ area}}{Q2 \text{ area}}$	$\frac{Q1 \text{ area}}{Q3 \text{ area}}$	$\frac{Q2 \text{ area}}{Q3 \text{ area}}$	Average	Standard deviation	Percentage of standard deviation	Similarity index (%)
1.00	0.86	0.86	0.91	0.0817	9.03	91.0

## Appendix D: References

- Allen, J. (2008). *More Tools for Teaching Content Literacy*. Stenhouse Publishers.
- Angelo, T. (1991). *Classroom Research: Early Lessons from Success. New Directions for Teaching and Learning*. Jossey-Bass.  
<https://doi.org/10.1002/tl.37219914603>
- Angelo, T., & Cross, P. (1993). *Classroom Assessment Techniques. A Handbook for College Teachers*. Jossey-Bass.
- Angra, A., & Gardner, S. M. (2018). The Graph Rubric: Development of a Teaching, Learning, and Research Tool. *CBE-Life Sciences Education, 17*, 1-18.  
<https://doi.org/10.1187/cbe.18-01-0007>
- Baldwin, S. J., & Ching, Y. H. (2019). An Online Course Design Checklist: Development and Users' Perceptions. *Journal of Computing in Higher Education, 31*, 156-172. <https://doi.org/10.1007/s12528-018-9199-8>
- Bannert, M., & Mengelkamp, C. (2008). Assessment of Metacognitive Skills by Means of Instruction to Think Aloud and Reflect When Prompted. Does the Verbalisation Method Affect Learning? *Metacognition and Learning, 3*, 39-58.  
<https://doi.org/10.1007/s11409-007-9009-6>
- Bartels, H., & Kulgemeyer, C. (2019). Explaining Physics: An Online Test for Self-Assessment and Instructor Training. *European Journal of Physics, 40*, 1-11.  
<https://doi.org/10.1088/1361-6404/aaeb5e>
- Berthoff, A. (1982). *Forming, Thinking, Writing: The Composing Imagination*. Boyton/Cook.
- Bresnock, A. E., Graves, P. E., & White, N. (1989). Multiple-Choice Testing—Question and Response Position. *Journal of Economic Education, 20*, 239-245.  
<https://doi.org/10.1080/00220485.1989.10844626>
- Butler, A. C. (2018). Multiple-Choice Testing in Education: Are the Best Practices for Assessment Also Good for Learning? *Journal of Applied Research in Memory and Cognition, 7*, 323-331.  
<https://doi.org/10.1016/j.jarmac.2018.07.002>
- Caruso, A. E., Hobart, T. R., Botash, A. S., & Germain, L. J. (2019). Can a Checklist Ameliorate Implicit Bias in Medical Education? *Medical Education, 53*, 1.  
<https://doi.org/10.1111/medu.13840>
- Chamizo, J. A. (1995). Mapas conceptuales en la enseñanza y la evaluación de la química. *Educación Química, 6*, 118-124.  
<https://doi.org/10.22201/fq.18708404e.1995.2.66719>
- Chamizo, J. A. (1996). Evaluación de los aprendizajes en química. Segunda parte: Registros de aprendizaje, asociación de palabras y portafolios. *Educación Química, 7*, 86-89. <https://doi.org/10.22201/fq.18708404e.1996.2.66671>
- Cheng, M. W. T., & Chan, C. K. Y. (2019). An Experimental Test: Using Rubrics for Reflective Writing to Develop Reflection. *Studies in Educational Evaluation, 61*, 176-182. <https://doi.org/10.1016/j.stueduc.2019.04.001>
- Cockett, A., & Jackson, C. (2018). The Use of Assessment Rubrics to Enhance



- Feedback in Higher Education: An Integrative Literature Review. *Nurse Education Today*, 69, 8-11. <https://doi.org/10.1016/j.nedt.2018.06.022>
- Cowan, J., & George, J. (2004). *A Handbook of Techniques for Formative Evaluation: Mapping the Student's Learning Experience*. Routledge Falmer.
- da Silva, A. C. (2019). A Conceptual Questionnaire on Radiations: Formulation Process and Analysis of Distractors. *Gondola Enseñanza y Aprendizaje de Las Ciencias*, 14, 63-79. <https://doi.org/10.14483/23464712.13113>
- del Pozo, J. A. (2012). *Competencias profesionales. Herramientas de evaluación: El portafolios, la rúbrica y las pruebas situacionales*. Narcea.
- Dodd, D. K., & Leal, L. (1988). Answer Justification—Removing the Trick from Multiple-Choice Questions. *Teaching of Psychology*, 15, 37-38. [https://doi.org/10.1207/s15328023top1501\\_8](https://doi.org/10.1207/s15328023top1501_8)
- Dreimuller, N., Schenkel, S., Stoll, M., Koch, C., Lieb, K., & Juenger, J. (2019). Development of a Checklist for Evaluating Psychiatric Reports. *BMC Medical Education*, 19, Article No. 121. <https://doi.org/10.1186/s12909-019-1559-1>
- Fosado, R. E., Martínez, A., Hernández, N., & Ávila, R. (2018). The e-Portfolio as a Transversal Tool for Planning and Evaluating Autonomous Learning for Sustainable Development. *Revista Iberoamericana Para La Investigación y El Desarrollo Educativo*, 8, 194-215. <https://doi.org/10.23913/ride.v8i16.338>
- Frederiksen, N., Saunders, D. R., & Wand, B. (1957). The In-Basket Test. *Psychological Monographs: General and Applied*, 71, 1-28. <https://doi.org/10.1037/h0093706>
- Hatzikiriakou, K., & Metallidou, P. (2009). Teaching Deductive Reasoning to Pre-Service Teachers: Promises and Constraints. *International Journal of Science and Mathematics Education*, 7, 81-101. <https://doi.org/10.1007/s10763-007-9113-8>
- Herrington, D. G., & Sweeder, R. D. (2018). Using Text Messages to Encourage Meaningful Self-Assessment outside of the Classroom. *Journal of Chemical Education*, 95, 2148-2154. <https://doi.org/10.1021/acs.jchemed.8b00361>
- Jasdilla, L., Fitria, Y., & Sopandi, W. (2019). Predict Observe Explain (POE) Strategy toward Mental Model of Primary Students. *Journal of Physics: Conference Series*, 1157, 1-6. <https://doi.org/10.1088/1742-6596/1157/2/022043>
- Kearney, M. (2004). Classroom Use of Multimedia-Supported Predict-Observe-Explain Tasks in a Social Constructivist Learning Environment. *Research in Science Education*, 34, 427-453. <https://doi.org/10.1007/s11165-004-8795-y>
- Keeley, P. (2015). *Science Formative Assessment: 75 Practical Strategies for Linking Assessment, Instruction, and Learning* (2nd ed.). Corwin Press.
- Kerr, W. C., & Macosko, J. C. (2011). Thermodynamic Venn Diagrams: Sorting out Forces, Fluxes, and Legendre Transforms. *American Journal of Physics*, 79, 950-953. <https://doi.org/10.1119/1.3599177>
- Lancaster, E. L. (2007). Questions and Answers (Encouraging Students to Sing and Count out Loud). *Clavier*, 46, 48.
- Latifah, S., Irwandani, I., Saregar, I., Diani, R., Fiani, O., Widayanti, W., & Deta,

- U. A. (2019). How the Predict-Observe-Explain (POE) Learning Strategy Remediate Students' Misconception on Temperature and Heat Materials? *Journal of Physics: Conference Series*, 1171, 1-6.  
<https://doi.org/10.1088/1742-6596/1171/1/012051>
- Lyons, H. (1999). How Portfolios Can Shape Emerging Practice. *Educational Leadership*, 56, 63-65.
- Metcalfe, J., & Xu, J. (2018). Learning from One's Own Errors and Those of Others. *Bulletin & Review*, 25, 402-408.  
<https://doi.org/10.3758/s13423-017-1287-7>
- Ogle, D. (1986). K-M-L: A Teaching Model That Develops Active Reading of Expository Text. *Reading Teacher*, 39, 564-570.  
<https://doi.org/10.1598/RT.39.6.11>
- Pendley, B. D., Bretz, R., & Novak, J. (1994). Concept Maps as a Tool to Assess Learning in Chemistry. *Journal of Chemical Education*, 71, 9-15.  
<https://doi.org/10.1021/ed071p9>
- Quintanilla, M., Labarrere, A., Díaz, L., Camacho, J., Cuellar, L., & Ravanal, E. (2008). El inventario de ideas previas (KPSI) como un instrumento de regulación de los procesos de desarrollo profesional de docentes de ciencias naturales en ejercicio. *Boletín de Investigación Educativa*, 22, 97-114.
- Robles, J. (2019). Rotation Self-Evaluation Checklist (REC): A Preceptors' Perspective. *Currents in Pharmacy Teaching and Learning*, 11, 258-263.  
<https://doi.org/10.1016/j.cptl.2018.12.005>
- Schippmann, J. S., Prien, E. P., & Katz, J. A. (1990). Reliability and Validity of In-Basket Performance Measures. *Personnel Psychology*, 43, 837-859.  
<https://doi.org/10.1111/j.1744-6570.1990.tb00685.x>
- Sridharan, B., & Boud, D. (2019). The Effects of Peer Judgements on Teamwork and Self-Assessment Ability in Collaborative Group Work. *Assessment & Evaluation in Higher Education*, 44, 894-909.  
<https://doi.org/10.1080/02602938.2018.1545898>
- Sutton, C. (1980). The Learner's Prior Knowledge: A Critical Review of Techniques for Probing Its Organization. *European Journal of Science Education*, 2, 107-120. <https://doi.org/10.1080/0140528800020202>
- Tamir, P., & Lunetta, V. N. (1978). An Analysis of Laboratory Activities in BSCS. Yellow Version. *American Biology Teacher*, 40, 426-428.  
<https://doi.org/10.2307/4446325>
- To, J., & Panadero, E. (2019). Peer Assessment Effects on the Self-Assessment Process of First-Year Undergraduates. *Assessment & Evaluation in Higher Education*, 44, 920-932. <https://doi.org/10.1080/02602938.2018.1548559>
- Tobajas, C., Molina, C. B., Quintanilla, A., Alonso-Morales, N., & Casas, J. A. (2019). Development and Application of Scoring Rubrics for Evaluating Students' Competencies and Learning Outcomes in Chemical Engineering Experimental Courses. *Education for Chemical Engineers*, 26, 80-88.  
<https://doi.org/10.1016/j.ece.2018.11.006>
- Treagust, D. F., & Chong, W. L. (1995). A Predict-Observe-Explain Teaching

- Sequence for Learning about Students' Understanding of Heat and Expansion Liquids. *Australian Science Teachers Journal*, 41, 68-71.
- Wardrop, J. (2012). Speaking out Loud: Muslim Women, Indian Delights and Culinary Practices in eThekweni/Durban. *Social Dynamics—A Journal of African Studies*, 38, 221-236. <https://doi.org/10.1080/02533952.2012.717209>
- Witchel, H. T., Guppy, J. H., & Smith, C. F. (2018). The Self-Assessment Dilemma: An Open-Source, Ethical Method Using Matlab to Formulate Multiple-Choice Quiz Questions for Online Reinforcement. *Advances in Physiology Education*, 4, 697-703. <https://doi.org/10.1152/advan.00081.2018>
- Wygoda, L., & Tague, R. (1995). Performance-Bases Chemistry: Developing Assessment Strategies in High School Chemistry. *Journal of Chemical Education*, 72, 909-911. <https://doi.org/10.1021/ed072p909>
- Zakaluk, B., Samuels, J. S., & Taylor, B. (1986). A Simple Technique for Estimating Prior Knowledge: Word Association. *Journal of Reading*, 30, 56-60.