The way students' internalize assessment criteria on inquiry reports

Inês Bruno[a,*], Leonor Santos[b], Nilza Costa[c]

[a] *Escola Secundária José Saramago, Avenida Cidade de Leiman, 2640-470 Mafra, Portugal*
[b] *Instituto de Educação e Centro de Investigação em Educação, Universidade de Lisboa, Alameda da Universidade, 1649-013 Lisboa, Portugal*
[c] *Departamento de Educação e Psicologia, Centro de Investigação em Didática e Tecnologia na Formação de Formadores, Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal*

A B S T R A C T

This paper, based on two case studies, presents an interpretative research on the processes used by high school students to internalize the assessment criteria of Physics and Chemistry inquiry reports. Findings support that understanding the assessment criteria is complex, mainly because of its terminology. However the discussion of exemplars can have a crucial role in this process. In order to operationalize the assessment criteria, students used strategies stemming from the social context (teacher, peers and other didactic sources) as well as from their individual experience (errors made, engagement in critical thinking and the creation of a favorable environment). However it also showed that some differences in operationalizing the assessment criteria were related to different students' profile. The results point to a multi-strategy pedagogical approach to enhance students' internalization of assessment criteria. Nevertheless reducing the tension between teachers' expectations and students' own standards of quality showed to be a complex process.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Modern societies require individuals, and therefore students in particular, to be able to continuously acquire new knowledge and competences. It is very unlikely that students succeed unless they reflect on themselves, on what and how they are doing (Dann, 2002). Thus, students need to have an active and constructive role in the learning process (Nunziati, 1990; Zimmerman, 2002). In this sense, self-regulation is an essential competence because it includes "the mechanisms of orientation, control, and adjustment of the cognitive, affective and social activities that promote the transformation of students' competencies" (Allal, 2007, p. 9). The assessment criteria play a crucial role in this process because they can guide students to achieve the objectives of a task (Nunziati, 1990).

However, each individual establishes his/her own standards, criteria or self-representations, so it is essential to use a pedagogical approach that fosters the internalization of assessment criteria, as suggested by Andrade and Du (2007), Santos and Gomes (2006) and Semana and Santos (2010). According to Vygotsky (1934/1978), the internalization of the assessment criteria means that students integrate them in their own knowledge. In other words, it implies that they: i) interpret the criteria, understand their meaning, reach a common understanding and appropriate the teacher's expectations (Vial, 2012), and ii) operationalize the criteria, in other words putting them into practice, because "criteria assume meaning only when used" (Woolf, 2004, p. 488).

In this article we assume that understanding an assessment criterion presupposes that the student gives the proper meaning to all the content of the criterion statement (matching the teacher's interpretation) and can explain it in his/her own words. Operationalization of a criterion means that students use it in specific situations.

According to several studies (e.g. Kirby & Downs, 2007; Rust, Price, & O'Donovan, 2003; Tillema, 2014), informing students about their assessment criteria is usually insufficient for internalization to occur. In fact, the internalization is a complex process (Andrade & Du, 2007; Clark, 2012). For this reason, and according to several studies (e.g. Tierney, 2013), the internalization process needs to be deeply investigated. The present paper aims to fill a gap in studies focusing on this process. The research was undertaken in order to understand how high school students internalize assessment criteria. It was guided by the following research questions:

* Corresponding author.
   E-mail addresses: inesbruno@gmail.com (I. Bruno), mlsantos@ie.ul.pt (L. Santos), nilzacosta@ua.pt (N. Costa).

1. How do high school students understand the assessment criteria of Physics and Chemistry inquiry reports?
2. Which strategies do students use to support the operationalization of assessment criteria when they develop Physics and Chemistry inquiry reports?
3. What is the role of the pedagogical approach in the internalization of the assessment criteria?

2. Conceptual framework

Self-regulation can be defined as "a multilevel, multicomponent process that targets affect, cognitions, and actions, as well as features of the environment for modulation in the service of one's goals" (Boekaerts, Maes, & Karoly, 2005, p. 150). Therefore, it assumes that students are actively involved in the learning process. According to the model proposed by Zimmerman (2002), self-regulation includes three phases: the forethought (where students define specific goals and plan strategies), the performance (where students monitor their performance) and the self-reflection phase (where students reflect on the methods, knowledge acquired and the importance of the process they used to achieve the goals).

Self-regulation fosters an awareness of when and how students learn, or not (Black, Harrison, Lee, Marshall, & Wiliam, 2002; Swaffield, 2011). It implies they construct their own learning path also through the understanding of assessment criteria, self-assessment and self-correction (Nunziati, 1990). This is essential for understanding how successful their work was, as well as for identifying and correcting their mistakes and errors (Dann, 2002; Papaleoutiou-Louca, 2003; Zimmerman, 2002). But for this to happen, it is necessary to make explicit what usually is implicit, promoting the understanding of assessment criteria.

Assessment criteria are representations of various aspects of the task (Nunziati, 1990); therefore its significance must be interpreted so that students understand what is actually requested (Vial, 2012). This process can occur unilaterally or bilaterally, with the teacher either informing or negotiating assessment criteria with students, respectively. In fact, the usefulness of the assessment criteria depends not only on their understanding, but also on the degree of their acceptance (Hadji, 1989). It is crucial that, from the point of view of the students, they make sense, that is, they must be accepted as important and legitimate. So, the bilateral process has the advantage that responsibility in the assessment process is shared with the students (Gipps, 1999). As such, an open and constructive dialog between teacher and students can help them to understand and consider the task requirements (Woolf, 2004).

Although assessment criteria often refer to students' tasks, they are abstract (e.g. rigor, clarity), thus it is essential that students are aware about the qualities their work should have. A useful way of doing this is by using rubrics (Brookhart, 2013). These are documents that demonstrate the characteristics that are expected from a particular task, indicating what leads to a high mark and describing levels of quality, for example from excellent to poor (Andrade, 2000). Rubrics can be analytical, when the performance is described on each criterion separately, or holistic, if the criteria are treated together. They also can be generic when a general performance is described and in this way they can be used with a family of similar tasks or be task-specific if the description of the performance is specific to a single task. According to Brookhart (2013), using analytic rubrics is more advantageous for formative assessment because it can help students to identify what aspects of their work need attention. The highest level, which indicates the characteristics of an excellent work, makes students to engage in a process of constructive learning based on self-regulation (Hafner & Hafner, 2003).

As several studies point out (Andrade, 2001; Green & Bowser, 2006; Santos & Semana, 2015), it is necessary to explore rubrics with students in order to reach a shared understanding. A clear explanation of what is expected helps students to understand the more unclear or less obvious items (which would otherwise be ignored), and overcome the vague, imprecise and subjective parts of some indicators. In the case of more abstract criteria, it is useful to give concrete examples for discussion, such as previous assignments of the students or other students' assignments, as suggested by Andrade, Du, and Mycek (2010), as well as by Hendry, Armstrong, and Bromberger (2012). Sometimes students need to see how assessment criteria can be applied. This is a useful way to help them understand and operationalize those criteria, or in other words, to internalize them. However, it is important to dissuade students from using mechanistic strategies or thinking that these examples are standard (Norton, 2004).

Feedback can also be used to promote the understanding of assessment criteria because ambiguities and misunderstandings may be diminished through this process (Taras, 2003). According to Nicol and Macfarlane-Dick (2006) and Nicol (2010), feedback, understood as a dialog process, may: i) help clarify what constitutes a good performance, which is crucial for students to set goals and guide them in the process of self-regulation; ii) encourage students to identify the criteria and standards that apply to their work and make judgments based on such standards; iii) give high quality information to students about their learning. This information should help them improve their own performance and correct possible mistakes. It should lead students to act so that discrepancy between intentions and results is diminished (Black, Harrison, Lee, Marshall, & Wiliam, 2003). In this way feedback can also foster students' operationalization of the assessment criteria. Therefore, it must be strictly related to the objectives, standards and criteria and students have to understand this relationship (Wharton, 2003).

Feedback also gives students the opportunity to engage in other ways of looking at the task, so that other paths can be considered (Taras, 2001). However, in the long run, the amount and detail of feedback should be minimized in order to promote an autonomous learning and the development of skills that enable the students to analyze their work independently (Black & Wiliam, 1998; Crisp, 2012; Santos & Pinto, 2009; Swaffield, 2011).

The dialog between peers can also be beneficial for the internalization of assessment criteria to occur (Price, Handley, Millar, & den Outer, 2007) since the language used by peers is easier to understand and the contact with other points of view and alternative strategies help students to review and build new apprenticeships (Black et al., 2003).

Previous studies considered the potential of enhancing the internalization of assessment criteria for increasing learning capacity. Students can identify more effectively the strengths and weaknesses of their processes and products, have better performances, be more motivated and less anxious and develop skills inherent to self-regulated learning (Andrade & Du, 2007; Dann, 2002; Rust et al., 2003).

3. Methodology

This study examines the processes used in the development of inquiry reports guided by assessment criteria. Therefore, and taking into account that the research questions are designed to address the understanding of the process of internalization of those criteria, we have chosen an interpretive paradigm (Lichtman, 2006). A qualitative case study was designed for this investigation. According to Merriam (1988), the qualitative case study 'is the ideal design for understanding and interpreting observations of educational phenomena' (p. 2). As a result, it was possible to

understand how high school students internalize the assessment criteria of Physics and Chemistry inquiry reports.

3.1. Participants

The study took place in a school in Portugal, and involved one class of Chemistry and Physics, over two school years. In the first year, the first author of this paper was the teacher of the class, and in the second year, another teacher took over but worked closely with the first one on the pedagogical intervention developed in the study. These teachers had respectively six and twenty years of professional experience and both were familiar with the use of inquiry tasks in the classrooms, but had a limited experience in using pedagogical approaches to promote the internalization of assessment criteria. The students (15/16-years-old) were informed about the nature and objectives of the study. Consents were obtained from all students and their parents. Two students were selected for case studies: Sara and Gustavo (participants were given fictitious names to ensure confidentiality and anonymity). These particular students were chosen because, while they had a good ability to express themselves, they had difficulty in reflecting about their own work. In addition, they were of different gender, performance level (between average and good) and they were available to be interviewed.

3.2. Pedagogical context

Four inquiry tasks were implemented over the 10th and 11th grades of the same class (from November 2008 to January 2010). As preconized in the National Science Education Standards (National Research Council, 1996), these tasks involved making observations, examining sources of information to see what is already known, planning investigations, using tools to gather, analyze and interpret data, proposing answers, explanations and predictions, and communicating the results. These tasks can be classified as guided inquiry activities (Bell, Smetana, & Binns, 2005) given that the inquiry questions were introduced by the teacher but the answers were given by students after designing a procedure and analysing the collected data independently. The guidelines given to students' work in the third task is presented as an example (Appendix A).

Writing requires students "to reflect, consolidate, elaborate, reprocess concepts and ideas central to the topic, hypothesize, interpret, synthesize, and persuade, and hence develop higher-order thinking and the construction of a deeper understanding of science concepts" (Yore, Bisanz, & Hand, 2003). Therefore, students were requested to develop reports containing five sections: introduction, planning, procedure, results and conclusions. Three of the reports included the development of a second version considering the written feedback provided by the teacher.

Students were also requested a written reflection about each report that consisted in assigning a grade level to each section considering the levels of the rubric (Appendix B) and giving a justification (pointing out its strengths and weaknesses).

Each inquiry task was developed in a 135 minutes lesson (1st lesson). The reports and written reflections of the second and fourth tasks were done during the next two 90 minutes lessons (2nd and 3rd lessons) and the report of the third task as well as the revision of the first version of the report of the fourth task were done in the school library after the schoolwork.

The pedagogical approach used in this study to promote the internalization of assessment criteria included:

i) *the discussion of the assessment criteria and the use of an analytical rubric.* Before the first task, the teacher gave students some guidelines about the structure of the report that included some questions for each section, such as: "What conclusions did you reached? Which were the limitations of the strategies you used? Would you change the way you solved the problem? How? And why?" Based on these guidelines, students were asked to collaborate in the selection of the most relevant assessment criteria, describing the characteristics of a quality report. Later on, the teacher developed a rubric taking into account students' suggestions. This was delivered to the students when they started the first task. As suggested by Griffin (2009), some changes were made after the teacher/researcher had analyzed the first task reports and the class discussion (in the light of the rubric) that followed the delivery of the final version (where students analyzed the quality of their performance considering each assessment criterion and made some suggestions about the content of the rubric descriptors). This new rubric was used by students in the next phases of the study (Appendix B).

ii) *the written and oral feedback.* Both types of feedback were used in a constructive way. The comments, usually provided in the form of questions or hints by the teacher, aimed to give student information about their performance in relation to an assessment criterion in order to help them to identify the cause of the gap between the desired and the actual level, as well as to fill that gap and to understand how to successfully operationalize that assessment criterion (Black & Wiliam, 1998).

iii) *the discussion of an exemplar report.* During the discussion of the reports of the second task, an exemplar report was given to the students, as suggested by several authors, like Andrade, Du, and Mycek (2010). This exemplar report was the compilation of the best parts of the reports done by the students of this class. It was carefully analyzed with the students through an open dialog between them and the teacher. Regarding each assessment criterion, students were asked to: i) explain the meaning of some words in the rubric, whenever it seemed necessary (e.g. synthetic), and ii) do a reasoned judgment about the quality of its' operationalization in the provided example.

iv) *the dialog between peers.* Only the first task was developed individually for the purpose of gathering data to select the case studies (though interaction between the students was allowed). The other tasks as well as the respective reports and written reflections were done by groups of three or four students. These groups composition changed from task to task and the selected students, Sara and Gustavo, were always integrated into different groups. The interaction between students was systematically encouraged. The teacher generally gave feedback only when she was asked to and after group discussion about that issue.

3.3. Collection of data

Observation, interview and documental analysis were used for collecting data. Observations and interviews were both audio-taped and later fully transcribed. The first author made participant observation (Cohen, Manion, & Morrison, 2000) during the lessons in which students performed the tasks and developed the respective reports and written reflections.

Semi-structured interviews (Burns, 2000) were carried out with the two case-study participants after submission of the last version of each report and at the end of the study. These allowed students to reveal feelings, intentions, meanings and thoughts about a topic or situation (Lichtman, 2006). The questions sought to encourage the analysis of each section of the report in order to understand the processes used for internalizing the assessment criteria and the reasons behind the less successful items.

Documents provide access to thoughts, ideas and meanings expressed by their authors and may raise questions that can be taken up in subsequent moments of data collection (Lichtman, 2006; Yin, 2002). In this study the collected documentation

consisted mostly of student reports and written reflections. These assignments reflect a more or less successful internalization of the assessment criteria, so their analysis raised questions that guided further data collection. Examples of student assignments, as well as excerpts transcribed from their reports were translated from Portuguese. The excerpts of the reports are identified by the letter R followed by the number of the task and the number of the version (V).

3.4. Data analysis

Data analysis was conducted in two phases. The first, which was developed during the data collection, raised issues to be considered in subsequent data collection. The second, more thorough and organized, occurred after the fieldwork and aimed to answer the research questions of the study. The method of questioning and constant comparison (Strauss & Corbin, 1998) was applied, using categories defined both during the analysis process and on the basis of the theoretical framework.

As stated before, the internalization of the assessment criteria includes the understanding and the operationalization of the criteria in a specific situation (Vial, 2012; Woolf, 2004). Thus, the following categories were considered: i) *understanding* and ii) *strategies used in the operationalization of the assessment criteria*. Regarding the second category, some subcategories emerged from the analysis: *discussing with peers and asking for help*, *searching for information*, *learning with the errors*, *critical interaction with themselves* and *creating a favorable environment*. The role of the pedagogical approach to foster the internalization of the assessment criteria was studied in a transversal way, focusing on the discussion of the assessment criteria and the use of the rubric, the feedback, the use of an exemplar report and the dialog between peers.

4. Results

4.1. The case of Sara

Sara is a student with a good performance in the subject of Physics and Chemistry. Her final grade was fourteen out of twenty[1] on both 10th and 11th high school years. She is a hardworking student, committed to achieve high marks, but she often feels insecure about her performance. She often asked her peers and teachers for help and sometimes she preferred not performing the task rather than making mistakes: "I knew it more or less, but I didn't know if I was right, since I didn't want to fail, I didn't write anything" (1st Interview). At the beginning of the study, she had limited experience in carrying out inquiry tasks and developing lab reports as well as assessing her own assignments: "Usually teachers tell me what to do, help me and now I have to think and do it by myself. That is why it was difficult" (1st Interview).

4.1.1. Understanding the assessment criteria

In Sara's view, the initial discussion of the assessment criteria, which lead to the establishment of the rubric, helped her to understand what was required to do in each section of the report:

> Researcher—We have discussed the assessment criteria of the reports. Do you think this was helpful?
> Sara—Yes, because, as it often happens, when I'm doing an assignment, I'm not sure about what I need to include or not and in this way I understood what I should or should not do. Yes, I

think it helped. (1st Interview)

However, the evidence suggests that this initial discussion and the reading of the rubric, by itself, did not promote the understanding of some criteria, mainly those that included aspects or expressions she was not familiar with and the criteria involving the use of complex skills.

For example, one of the assessment criteria was "defining *clear and synthetic objectives* consistent with the task". Initially, it appears that Sara and her colleague were not able to give a proper meaning to the word *synthetic*:

> Student A—What is synthetic? Synthetically? Clear and simple, isn't it?
> Sara—Yes, I think so. (2nd Task; 3rd Lesson)

However, the analysis of the exemplar report appears to have promoted the understanding that the goals should be direct and concise. This idea is based on, firstly, the fact that on the second interview (which occurred after the discussion of the exemplar report), Sara pointed out that the objectives were one of the weaknesses of her report because they were not concise [sentences 1 and 2 below]; secondly, in the report of the third task (R3), she was able to formulate a synthetic goal and in the subsequent interview she described its characteristics [sentences 3 and 4]:

> 1. Sara—We should explain our goals explicitly and we were . . . Basically, instead of saying the essential, we were verbose, let's say.
> (. . .)
> 2. Sara—Then, with the report the teacher gave, which was an example (. . .) I noticed that they went straight to the point and didn't ramble about. They just say what it is and this was not quite what we have done. (2nd Interview)
> 3. Researcher—What is your understanding of a synthetic goal?
> 4. Sara—Not very long and not scattered throughout the text (. . .) said in a plainly way. (3nd Interview)

The objectives of this activity were to collect experimental data to know what was the most effective way to cool lemonade. (R3)

Regarding the *indication of alternative strategies to solve the problem*, Sara initially had some difficulty to understand what was expected. So, during the first task, the teacher gave her an example that could be solved using different separation techniques (something that was familiar to her because this subject had been previously studied). But, the discussion in the classroom concerning the different instruments that could be used to measure the volume of sand seemed to have led Sara to consider that these different options were different strategies to solve the problem. In the report, she identified several instruments and presented arguments to justify her choice: "To have a greater accuracy in the results I chose to use a measuring cylinder because the beaker has a low precision and the pipette is not suitable for measuring the volume of sand (a solid) because it is used to measure the volume of liquids" (R1_V2).

The evidence suggests that the discussion about the written reports (when the second version was delivered to the students) helped her to create a wider understanding of what an alternative strategy is. During the interview, that occurred the day after this class, Sara was able to criticize her assignment, identifying why alternative strategies were missing and why she previously thought they were present:

> Sara—I didn't wrote several strategies, I only wrote one (. . .) I thought it was the instrument that we used to measure the volume, thought it was the beaker, the pipette . . . and I did not know that it was expected a really different strategy, then I assessed myself with level 3, but it is not correct.

---

[1] In the Portuguese educational system grades range from 0 to 20, with 10 or higher being a "positive" passing mark.

Researcher—Okay, so here you've considered that presenting several instruments to measure meant to present different strategies, that's why you assessed yourself with level 3.
Sara—And because I also explained why, I thought I was doing ok. (1st Interview)

Another assessment criterion that was not initially understood by Sara was the *presentation of evidence to support the conclusions*. As the teacher drew attention to the importance of presenting facts that support the conclusions [sentence 3], Sara suggested to take this oral feedback into consideration, but she included the values of the controlled variables [sentence 4], instead of the findings that could justify her conclusions. Besides, the comment that was written in the first version of the report of the second task ("In these conditions, how many molecules/atoms do you have? You should always justify your conclusions based on your results") was not effective as well.

During the following interview, Sara revealed that she had not previously understood this assessment criterion [sentences 10–13] and could not find it on the rubric as well [sentences 8 and 9]:

1. Researcher—So, if we are in the same conditions of pressure, temperature and the volume remains constant . . .
2. Sara—It will always have the same number of atoms or molecules.
(. . .)
3. Researcher—Exactly. And now you should justify that conclusion presenting the findings of your experiment, your simulation.
(. . .)
4. Sara—You have to insert the value of pressure because you have to include your results.
5. Student A—Is that so?
6. Sara—Yes. (2nd Task; 1st Session)
7. Sara—We didn't always include our findings. We did that regarding some conclusions, but not all of them.
8. Researcher—But where can you find this orientation in the rubric?
(. . .)
9. Sara—It is not here [in the rubric]. It is about the conclusions, but it is not written. It is not here, but I know that we need to include it.
10. Researcher—So what does this mean: the conclusions are explicitly based on evidence?
11. Sara—(silence). I think it is: I draw conclusions of what we found with what really is . . . How can I say it? What we found with the reality, in this case with Avogadro's Law. If our results were in line with the Avogadro's Law.
12. Researcher—And what does this mean: that are explicitly [emphasis on explicitly] based on evidence?
13. Sara—Well, I don't know that. (2nd Interview)

In the third task, Sara explicitly referred to the findings to justify the conclusions, thus supporting the fact that the discussion and analysis of the exemplar report helped Sara to understand the teacher's expectations about this assessment criterion.

Finally, it was possible to realize that Sara did not understand what was meant about *explaining how the identified sources of error affect the results*. When asked about the presence of this item in the report, Sara initially considered it was present [sentences 1–4]. But then she considered that it was not present and that she had not understood what was required [sentences 5–10]:

1. Researcher—Did you explain how your results were affected by the errors you identified? How does the fact that the system is not isolated affect your results?
2. Sara—Yes.
3. Researcher—Where did you explain that?

4. Sara—It is here, for example . . .
(. . .)
5. Researcher—So, did you identify the errors and explain their effects on the results?
6. Sara—No. (. . .) Because we explained the errors but we didn't explain how these errors would . . .
7. Researcher—Affect the results?
8. Sara—Yes
9. Researcher—So, basically you have not yet realized what this means, is that it?
10. Sara—No, I guess I had not. (3th Interview)

4.1.2. Strategies used in the operationalization of the assessment criteria

Sara usually started by reading the higher level of quality of the rubric even though she did not seem to do an in-deep analysis of the descriptors: "when I read it . . . sometimes I don't pay much attention to the more difficult sentences. I don't understand, and then I pretend that it has already been done" (1st Interview).

After reading she used a number of strategies to operationalize some of the assessment criteria, such as *discussing ideas with peers* and *asking for help*. It seems that the interactions she established with peers and the teacher not only aimed at overcoming difficulties and identifying errors, but also as a form of validation:

Researcher—How did you overcome the difficulties that have arisen?
Sara—First, I compared the results with my colleagues, asked them how they were going to do. (. . .) Then they said what they were going to write, and I tried to do my best. (1st Interview)
Sara—Teacher, in the procedure, can we write: Firstly, secondly, thirdly? (1stTask; 2ndSession)
Sara—We have to indicate a name for the table. (. . .) We can write data record, right? Then you write: Table 1. (. . .) Ah, what do we write? Ah, measurements made?
Student B—Wait, we just measured the mass and the temperature. (3rd Task; 1st Session)
Sara—Don't we have to indicate the values? What do you think? (. . .) We add ice in the measure of . . .
Student B—No, we add 53 g of ice. (3rd Task; 1st Session)
Sara—I think there is something wrong, I am sure. Is this okay like this? (3rd Task; 1st Lesson)

The discussions and requests for assistance occurred in all sections of the report, but they were more frequent during the data analysis. Regarding this aspect, although the calculations usually had been correctly done, the reasoning was not always understood: "Now I know, but I didn't realize this before, why we said that the heat . . . ah, why we equal (. . .) I couldn't understand it well" (3th Interview).

In order to produce a more complete assignment and overcome difficulties, Sara also *searched for information* on the web, the school textbook, the former reports and the exemplar report. Usually, she incorporated the information almost verbatim (in particular with regard to the definitions):

Sara—We used the Internet to see the definition. We didn't want to write what we knew otherwise it would be poorly explained, but then when we tried to fit it . . .
Researcher—What definition did you search on the Internet?
Sara—The volume, the molar volume. Then, when we tried to add this, I think it ended up with some repetitions. But, then, if we had changed it, it would become even worse. (2nd Interview)

The fear of making mistakes and not being rigorous, as well as the lack of confidence in herself seems to have been one of the

reasons why Sara: i) fully copied some information available on the Internet: ""I think this part is very well done while this other one is not so good because we wrote our own thoughts on the paper. But it was not the same case here, we searched for and inserted the information (. . .) The way we express ourselves is not good at all" (2nd Interview); ii) transcribed the beginning of some sentences of the former reports: "We saw what we had written, how we had begun sentences" (2nd Interview); iii) used the structure of the procedure of the exemplar report, despite the procedure they had previously done also being correct: "We did [the procedure] in topics, but then as that exemplar report was done as a narrative, we thought it was more organized and decided to do the same. So we just had to switch to the narrative format" (3th Interview).

However, regarding the drawing of the experimental setup, the exemplar report was not used in a mechanistic way and seems to have been crucial for the operationalization of this criterion. In spite of the fact that the teacher had already acted to promote the internalization of this criterion, she was only able to operationalize it with the support of the exemplar report:

> Including a diagram of the simulation could help us comprehend the procedure, don't you think? (written feedback in the first version of the second task's report)
> Sara—Yes, I think that a diagram would have helped (. . .) We didn't think that we could do it as Luisa did [referring to the exemplar report] (. . .) And, because we were not able to do it as it should be done, it would be an invention, we preferred not to make a diagram. (2nd Interview)

While doing the report of the third task, Sara told her colleagues: "Look, we should do it [the diagram] the way it is here [the exemplar report]. I think it is very well done" (3rd Task; 1st Session) and then she managed to develop an adequate diagram (Fig. 1).

Learning with the errors was another way for Sara to operationalize the assessment criteria more successfully over time. In the second task, attention has been drawn (through oral feedback) for not rounding numbers when making calculations. From that moment on, she became careful about the accuracy of the data analysis, being aware of the reasons for this procedure:

> Student B—Are you going to introduce all digits?
> Sara—I will, otherwise we would have inaccurate results. (3rd Task; 1st Lesson)

Furthermore, concerning the scientific terminology, Sara did not repeat some of the mistakes on subsequent reports. For instance, instead of using incorrectly the word *weight* (pointed out by written feedback on the 1st report: "Does the scale measures weight?") she started to use the word *mass*. Her speech supports the idea that she tried to take into account the feedback provided in the former reports:
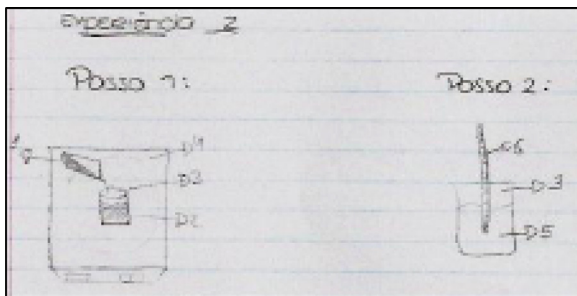

Fig. 1. Drawing of the experimental setup (R3).

Researcher—Were there some aspects that you paid more attention to in the report of this task because of the report you had done previously?
Sara—Yes, I remembered that I had not put the title on the tables. This was the first thing I remembered: to put the title on the tables. (2nd Interview)

Identifying the less successful aspects in the previous reports (through the feedback and supplemented with the discussion and analysis of them) also seems to have contributed to an improvement in the integration of concepts in the theoretical framework. For example, in the report of the fourth task, the concepts were explained in a more abstract way and did not constitute a direct description and explanation of the physical phenomena that occurred in practice. Expressions like "we have to calculate . . . " or "we multiply this value by . . . " no longer existed.

With regard to the planning, there was an improvement in the description of the calculations that would have to be made in order to answer the problem—it became clearer and more correct. For example, in the second task, Sara often wrote that they linked the variable x with y, but usually it was not explained how these variables were linked: "Explaining what we calculate is more difficult. (. . .) We were not able to explain what we had done." (2nd Interview). But in the two last reports, Sara included, in the description, the mathematical expressions, which facilitated its comprehension.

With respect to the conclusions, the sources of errors that were identified became more specific over time. While, in the report of the first task, Sara only wrote that accidental errors occurred, in the following reports the sources of errors became more specific to the task and more detailed:

> The largest temperature difference occurred in the ice experiment because more errors were committed. One of them is related with the fusion of the ice. Since the outside temperature is higher than the ice temperature (so the ice melts quickly) and we took some time to add the thermometer in the beaker and to measure the temperature, we have made an error. We also made other errors, such as measurements with the scale since a few drops of water may have been out of the beaker . . . (R3)

Lastly, it is important to add that sometimes taking into account the errors made previously was not enough. This strategy had to be complemented with others, e.g. discussing with peers. This was clear when Sara knew what to do (which aspects she should improve comparatively with previous reports), but did not know how to do it, as documented in the following:

> Another problem was the titles of the tables. We never knew what we should write. (2nd Interview)
> What will be the name of this table? (3rd Task; 3rd Lesson)

Another strategy used was the *critical interaction with herself*. Sara made suggestions to improve the completeness and accuracy of the assignments very often. She used to propose several ways to develop the work or to express her ideas until she got one that she liked:

> When we add ice to the lemonade it will give energy in the form of heat for a state change to happen, resulting in ice melting. This is all the change in enthalpy. No, only here, state change, occurs ice melting. Here we put enthalpy, change of enthalpy. (3rd Task; 2nd Lesson)

However Sara rarely adopted a critical attitude about the teacher's suggestions, even when these opposed her standards of quality. In the report of the fourth task, a written comment was made suggesting the aggregation of two tables: "Why don't you include these measures, which are also direct ones, in the table

above? (you can improve the organization)" (R4_V1). In the second version, only one table was presented to record the data, but Sara considered that this aggregation had worsened the organization of this section:

Sara—[after reading the written feedback] It is perfect the way it is!!! (4th Task; 1st Session)

Sara—We thought that it was better this way, with two tables. (. . .)

Researcher—So, you didn't agree with the comment, but you did it because the teacher wanted it!

Sara—Yeah, because the teacher wanted it, exactly. Because we thought we would be much . . . it would not be very well organized. The things would be on top of each other and then it would be more difficult to build the table. It wouldn't be as perfect as it was with two [separate tables]. (. . .) In our view it could be worse, but in the opinion of the teacher it could be better. (4th Interview)

Furthermore, Sara sought to *create a favorable environment* for the development of the report, suggesting doing some sections on an environment that allowed her to be more focused:

We should write all that we are saying, what we know, but it's complicated. Don't you want to go ahead to do something easier [instead of doing the theoretical framework]? We aren't very concentrated right now. (3rd Task; 2nd Lesson)

4.2. The case of Gustavo

Gustavo is a student with an average performance in the subject of Physics and Chemistry. His final grade was twelve and eleven out of twenty in years 10 and 11 of the high school curriculum, respectively. He is involved in several extracurricular activities and manages to get satisfactory results. He is confident in his abilities and reveals critical sense. At the beginning of the study, he had a limited experience in carrying out inquiry tasks and developing lab reports as well as assessing his own assignments: "I am not used to do these experiments (. . .) Now, I have learned to develop a well done report, which I didn't know how to do" (1st Interview).

4.2.1. Understanding the assessment criteria

Gustavo had considered the initial discussion about the assessment criteria beneficial, because it were the students who talked about what should be accomplished in order to develop a high quality work: "I think it was advantageous because we tried to present all the steps that needed to be done so, this way, it would be easier to make the report because we had already stated the steps previously" (1st Interview). However, he pointed out that the use of certain words in the rubric, such as "adequate" may not be enlightening if the student does not know its meaning in the context of the task. For instance, what is a suitable diagram to illustrate the procedure used to solve the problem: "On the last level it says: It includes an adequate diagram, but how do we know that . . . ahh, what is an adequate diagram? I do not know, maybe it could be a little bit more explicit" (3rd Interview). As Gustavo did not understand what was expected, he did not include the drawing of the experimental setup in the report of the first task.

In the first version of the report of the second task, it seems that he thought that the diagram should illustrate the calculations they had done to solve the problem instead of representing the experimental procedure (Fig. 2).

The exemplar report seems to have promoted the understanding of what was required, thus he developed a suitable drawing of the experimental setup in the report of the fourth task (Fig. 3): "I have difficulty to do the diagram. But after seeing the diagram of
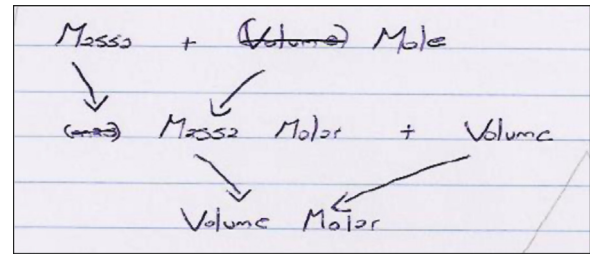

Fig. 2. Drawing of the experimental setup (R2_V1).

my colleagues [exemplar report] I thought: Ah, that's it!" (2nd Interview).

Regarding the assessment criteria *defining clear and synthetic objectives*, initially Gustavo did not understand the meaning of the word synthetic, but he only asked for a clarification after the first task, during the interview. Meanwhile, he assumed that the objectives should be presented in an easily understandable way [sentence 1], so the first ones included definitions [sentence 2]. However, the clarification provided during the class discussion (when the exemplar report was analyzed) seemed to have helped him to understand what was expected, and in the following task he proposed to write a short objective (not mixed up with the definition of the concepts) [sentence 5]:

1. What did the teacher mean by synthetic? I thought that it meant clear, in a way that it is well understood. Synthetic, I thought it was also something related to that as well: properly understandable. (1st Interview).
2. I tried to put myself in the shoes of a person who didn't know about this and whether, if I read the objectives to a person who did not know about the Avogadro's law, would this person be able to understand our objectives. (2nd Interview)
3. Student C—The objectives I present are consistent with the task and are clear and synthetic [reading the higher level of the rubric].
4. Student D—No, we have to say the approach, in other words, it is more or less the answer to the first question.
5. Gustavo—No, what is asked here, is only to say the objective. Then here [referring to the conceptual framework] we should say the rest. (3rd Task; 1st Session)

Another assessment criterion that was not initially understood by Gustavo was the *presentation of evidence to support the conclusions*. He thought that the conclusions should be drawn through a plain analysis of the results [sentence 1 below]. And, although the expectations about this criterion had been discussed with him, it seems that the teacher and Gustavo had not reached a common understanding because in the following reports he did
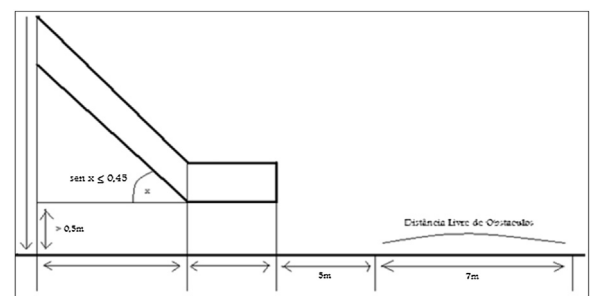

Fig. 3. Drawing of the experimental setup (R4_V1).

not make the results explicit on the conclusions; instead he considered that this assessment criterion had been successfully operationalized [sentences 2 and 3]:

1. It means that it is about things that we can look and immediately realize . . . we can immediately draw a conclusion (. . .). I thought that it was about being obvious. (2nd Interview)
2. The conclusions I take are explicitly based on the findings: well, I think we did that. (3rd Interview)
3. We tried to make explicit the results that we have found. (4th Interview)

A similar situation happened with the *explanation of how the identified sources of error affect the results*. Gustavo's analysis about the operationalization of this assessment criterion appeared that, at the beginning, he had not completely understood what was expected in order to apply it successfully, but at the end he showed that he understood it:

Researcher—Did you explain the effect of each source of error in your results?
Gustavo—[after reading his assignment] (. . .) I think we wrote what the sources of error did. I believe so.
Researcher—You wrote that the sources of error caused differences, but you didn't write that because an amount of water remained in the beaker, the final temperature was higher or lower than the predicted result.
Gustavo—(. . .) We didn't write how it did influenced . . . Exactly. I get it. I agree.
Researcher—Does it mean that you hadn't previously understood this part, what does this mean?
Gustavo—Well, I understood, but I didn't remember . . . we didn't remember to mention this, how it has influenced. We just wrote that it influenced and what influenced, but no how . . . Exactly, this is missing. (3rd Interview)

4.2.2. Strategies used in the operationalization of the assessment criteria

Gustavo usually started by reading the higher level of quality of the rubric and then he adopted several strategies to support the operationalization of the assessment criteria, such as *discussing ideas with peers* [sentences 1 to 5] and *asking for help* [sentence 6] to achieve accuracy and completeness of some parts of the report, especially those involving the data analysis. He sought not only to get a correct answer, but also to understand the reasoning behind it:

1. Gustavo—How do you calculate the molar mass?
(. . .)
2. Student D—Now you do: M equals four
3. Gustavo—Where did you get the four?
4. Student D—I saw it on the Periodic Table. Times one.
5.Gustavo—How do you know it is a one? (2nd Task; 1st Lesson)
6. Gustavo—Teacher, Gomes did it [the prediction of the results] this way. But I don't understand any of this. (3rd Task; 1st Lesson)

Another strategy used by Gustavo was *learning from errors*. For instance, the introduction of the reports became more complete over time (containing the definitions of the most important physical and chemical quantities):

Gustavo—I think a person also learns from mistakes, isn't it?
(. . .)
Researcher—Give me an example.
Gustavo—An example? Here, in the introduction, for instance, when you ask to explain what density is. The next time that I use

density I will have to mention it in the introduction, giving a detailed explanation.
Researcher—But, what if the report is not about the density, what if it is about another matter?
Gustavo—If it is about another subject I will also (. . .) I gave the example of the density. (1st Interview)

Concerning the scientific rigor, it seems that Gustavo also learned from errors because some mistakes made in the first reports (to which feedback had been provided) were not repeated in subsequent ones. For example, in the first version of the report of the first task, Gustavo wrote: "Then we divide the value of the mass by 10, in order to get the result in $g/cm^{3"}$ (R1_V1), without mentioning the physical quantity that he was calculating. Considering the written comment: "The result of what? What physical quantity are you calculating?" (R1_V1), he added, in the second version, the physical quantity that was being calculated (the density). Furthermore, in both versions of the report of the second task, Gustavo continued to indicate the physical quantity instead of its unit: 'to calculate the molar mass', "calculate the molar volume" (R2_V1), "calculate the number of particles" (R2_V2).

The organization of the tables improved as well. After a written comment being made as feedback to the first version of the report of the first task about the units that were registered in each row ("The advantage of indicating the units here (in the column heading) is not having to repeat it"), Gustavo began to show the units only in the title of the columns in the following reports.

Gustavo also appealed to a *critical interaction with himself* while he was developing the reports. Occasionally, he made suggestions and changes in order to improve the scientific rigor, the writing and the analytical accuracy:

This is the weight of the balloon. The weight of the balloon is equal to 96 point 753 (. . .) Wait, what weight?! This is the mass, not the weight. (2nd Task; 1st Lesson)
Because I started to do all the maths, but the numbers of the results were a bit large. So, [I thought] what is going on here? It is impossible to have a slide measuring 100 m. I saw that I had forgotten to put in the square on the velocity in the kinetic energy expression. (. . .) I was wrong many times. (4th Task; 3rd Lesson)

Gustavo also used his reasoning to define some words in the conceptual framework (and did not search for information): "I used my own words" (. . .) "Based on the name, I tried to . . . " (1st Interview).
Besides, in Gustavo performance sometimes prevailed the self-imposed standards, which means that, despite understanding the assessment criteria, Gustavo decided not to take into account some of them because it made more sense to him to do the assignment in a different way. This happened, for instance, regarding the presentation separately of the data collection and the data analysis, as well as, the synthesis of the objectives:

On the one hand, it is well done, but on the other, it is not well organized, perhaps because all the collected data should be together. But here, as we are showing some data and then doing the calculation . . . If we had only presented the calculations [separately], people could think: Okay, this is here, but where did this data come from? (. . .). I think it's better this way, because it is sequentially done . . . but it is not so scientifically correct. (3rd Interview)
Researcher—Synthetic means that it is something short and direct.
Gustavo—Exactly.
Researcher—For instance, regarding the horizontal displacement wouldn't you have the opportunity to explain, later on, in

the conceptual framework, what is it and how you calculate it? Did the definition of horizontal displacement have to be included in your objective?

Gustavo—It didn't have to, but we thought it would be better if it was there because, for instance, if the person who is going to read, doesn't know what it is (. . .) he/she keeps reading but is still thinking about what it is. However, if we include it right away, this person will know immediately what it is. (. . .) In my opinion, it is more correct to write the definition immediately after the word. (4th Interview)

In addition to the previous strategies, Gustavo tried to *create a favorable environment* to the development of the report. He suggested to do or to revise some parts of it, the ones that he considered more difficult or thought that were not as well done as he desired, in a quieter environment and by taking more time:

The strategy we have chosen was to calculate the energy given by the lemonade, as this is equal to the energy absorbed by the cold water or the ice, isn't it? I don't think this is properly explained. I have to see this at home. You have to send me this, Duarte. (. . .) Now I don't know what else to write. (3rd Task; 1st Lesson)

5. Discussion

In the two case study, students did not initially understand the totality of the assessment criteria, apparently because of the terminology that was used in the rubric and/or its high cognitive demand. Although an initial discussion about the assessment criteria happened in the classroom, the rubric descriptors were not written collaboratively with the students, and the terminology used was not always understood by them (e.g. presentation of evidence to support the conclusions). The results show that both students were not sure about the meaning of some words (e.g. synthetic), but they usually did not ask for clarifications. Moreover, in spite of the feedback that had been given regarding some criteria, Sara did not succeed (although she had shown intention to consider it). The absence of a common understanding about what was expected seems to be the reason for that. So, this study, like that of Clark (2012), suggests that one of the biggest difficulties in order to reach a common understanding is translating assessment criteria into a language that students can understand.

For both students, the analysis of the exemplar report seemed to have a crucial role in the process of reaching a common understanding, as previously suggested by Andrade, Du, and Mycek (2010), as well as Hendry et al. (2012). Seeing the operationalization of those criteria seems to contribute to the students understanding of the teacher's expectations.

In order to operationalize the assessment criteria, students used some strategies: discussing with peers and searching for help, collecting information from available sources, learning from mistakes, critical interaction with themselves and creating a favorable environment, although some differences were found. Sara, who is insecure, was not usually easily satisfied with her suggestions, so she frequently revised them. Furthermore, she often asked teachers and her peers for help in order to validate her ideas, and she searched for information. Wanting to achieve high marks but seeming to be afraid of taking risks (to fail), Sara sometimes transcribed the information with "guaranteed" quality. Gustavo, who is more confident and not as committed to achieve high marks as Sara, rarely searched for information and instead used his reasoning to accomplish the assignments.

Another important difference is that the primary concern of Sara seems to be to reach the teacher's expectations, even if she

does not completely understand or agree with the teacher's suggestions. However, Gustavo appears to be concerned with understanding the reasoning behind the criteria. His attitude shows that his work must make sense to him. This seems to justify, the tensions between the self-imposed standards and the assessment criteria that sometimes emerged throughout the study, as previously found by Andrade and Du (2007), Santos and Gomes (2006), and Semana and Santos (2010). Gustavo, seemed to need to identify a clear advantage in developing the assignments in a way that opposed his own quality criteria, otherwise he would consider his self-imposed standards (giving arguments for this procedure). However, Sara seemed to try to understand the "rules of the game" (Hadji, 1989) and to act in accordance with them. This is an aspect that several authors (e.g. Nicol & Macfarlane-Dick, 2006) have referred to—assessment criteria must be considered important by students in order to prevent them from acting in a certain way just to please the teacher. So, the tension between the self-imposed standards and the assessment criteria and the way this tension decreases reflects a complex process of internalization of assessment criteria (Andrade & Du, 2007).

In the light of the above, it seems that the way students operationalize the assessment criteria is more related with the students profile than with their average performance in the subject. Furthermore, there was a progress regarding the operationalization of most of the assessment criteria. However, some difficulties remained about some of them, which suggests that certain assessment criteria are more difficult to internalize than others.

Regarding the role of the pedagogical approach used for the internalization of the assessment criteria, the use of the rubric was important to guide the students, since they usually started by reading the highest-level descriptor of each section when they were doing the report. The use of former reports (with the written feedback) was conducive of learning from mistakes as previously reported (Santos & Gomes, 2006). This study, as well as that of Andrade, Du, and Mycek (2010), also suggests that the analysis of the exemplar report helped students to understand what was required from them. But one of the students (Sara) used the exemplar, as well as the former reports, to copy the layout and terminology (Hendry et al., 2012). The fact that this disadvantage was only evident in one of the case studies suggests that a mechanistic approach is more likely to arise when students are afraid of making mistakes and therefore look for models to copy.

The interaction between students helped mainly the overcome of difficulties in operationalizing the assessment criteria. Feedback, besides that, also favoured the identification of mistakes and omissions and the production of an improved version of the original assignment.

6. Concluding remarks

This study emphasises that understanding the assessment criteria is a complex process mainly because the terminology and the logic behind the subject and the teacher are usually different from the students (Nunziati, 1990; Vial, 2012). In order to operationalize the assessment criteria, students used strategies stemming from the social context (teacher, peers and other didactic sources) as well as from their individual experience (errors made, engagement in critical thinking and creating a favorable environment). The ways and frequency of how these strategies were used were variable and related to the students' profile, evidencing the relation between different components of self-regulation—cognitive, affective and social.

Regarding the role of the pedagogical approach in the internalization of the assessment criteria, it seems that the exemplar report was crucial for the understanding of the assessment criteria that had not been previously understood (in the initial discussion). But all teachers' interventions (discussing the assessment criteria, using the rubric, giving feedback, discussing an exemplar report and promoting dialog between peers) seem to have played a role in the operationalization of the assessment criteria (guiding the students, helping them to overcome difficulties and to learn from mistakes). However, notwithstanding the pedagogical approach been used to promote the internalization of the assessment criteria, this process proved to be complex (Andrade & Du, 2007), requiring continuity over time, even for high school students (with extensive experience in the school).

It should be highlighted that in this study researchers managed to guarantee the application of essential conditions to promote learning through assessment: an environment where students felt that mistakes are learning opportunities, using complex cognitive tasks and transparency of assessment criteria (Segers, Dochy, & Cascallar, 2003). Researchers intentionally sought to promote favorable conditions for the development of self-regulation (Wiliam, 2011), stimulating higher-level cognitive learning.

One of the limitations of this study was that, despite the awareness that the participation on this study would not influence their grades, students may have not shared all of their perceptions, emotions and opinions because the researcher (and interviewer) had been their teacher on the first year of collecting data. However, we consider that this double function (teaching and researching) did not have repercussions for the validity and reliability of the findings because the interactions between the teacher/researcher and students were based on confidence. We strongly believe that students felt comfortable sharing their feelings.

Some of the terminology used in the assessment criteria was difficult to understand, and even though the teacher and the students seemed to have reached a common understanding, the progress regarding the operationalization of those criteria had not always been consistent over time. Therefore, we suggest that further research should include analysis of the potentialities of writing the rubric's descriptor collaboratively with the students or asking them to: i) re-write the descriptors (using a familiar terminology); ii) take some notes after the discussion about them or; iii) underline the place where each assessment criteria is applied in the exemplar report and highlight the characteristics that show the quality of its operationalization.

Finally, as the methodology used in this study does not enable statistical generalization of the results, further research should extend this work to students of other years and levels, with different characteristics, and include additional tools of learning/assessment, as well as additional subjects. Nonetheless, the in-depth analysis of the two cases contributed to deepen the knowledge on students' internalization of assessment criteria.

Acknowledgments

Appendix A. —Students' Guidelines for the Third Task

Summer is coming and a fresh lemonade would be great! Have you ever thought about what is more efficient for cooling the lemonade: liquid water at $0\,°C$ or ice at $0\,°C$?

To get an answer to this question you have water at room temperature instead of lemonade and also water and ice at about $0\,°C$.

1. Which processes (adding water at $0\,°C$ or ice at the same temperature) will be more efficient for cooling lemonade? Justify your prediction.
2. Plan an experimental work in order to answer this problem.
3. Predict what will be the equilibrium temperature of water after adding some mass of water at about $0\,°C$. Justify your prediction.
4. Predict what will be the equilibrium temperature of water after adding some mass of ice at about $0\,°C$. Justify your prediction.
5. Perform the experimental work that was planned and discussed with the teacher. Note: You must break the ice and put it in a glass with water in order for the temperature of the inner ice to homogenize at $0\,°C$.
6. Compare the results with the initial predictions and discuss it.
7. What changes do you suggest to cool the water more efficiently?
8. How would you answer now to the question that guided your work?

Appendix B.

See Table B1.

Table B1
Rubric for Inquiry Reports.

Section	Level Items	1	2	3	4
Introduction	Objectives	The objectives I present are inconsistent with the task	The objectives I present are consistent with the task but are not clear	The objectives I present are consistent with the task and are clear	The objectives I present are consistent with the task and are clear and synthetic
	Scientific concepts related to the experiment	I refer to the scientific concepts related to the experiment, but don't explain them	I refer to the scientific concepts related to the experiment and explain them partially or give shallow explanations	I deeply explain all the scientific concepts related to the experiment	I deeply explain all the scientific concepts related to the experiment and justify their importance to the task
Planning	Strategy	The strategy I choose is not adequate to the objectives	The strategy I choose is adequate to the objectives		
		I present the strategy I choose with few details	I present the strategy I choose with a detailed description	I present the strategy I choose with a detailed description and explain my choice. I give justification for each step of the task.	I present the strategy I choose with a detailed description and explain my choice and the reasons I give up others. I give justification for each step of the task.
	Expected results	I outline the expected results, but do not provide logical reasoning for it	I outline the expected results and provide logical reasoning for it		
Procedure	Task development	I describe only a few steps of the task	I describe most of the steps of the task	I describe all the steps of the task	
	Drawing of the experimental setup	I illustrate the procedure but the diagram is not completely adequate	I illustrate the procedure with an adequate diagram	I illustrate the procedure with an adequate diagram, correctly identified with title and labels	
Results	Data collection	I present some collected data (measurements/ observations)	I present most of the collected data (measurements/ observations)	I present all the collected data (measurements/observations)	
		Data is reported with frequent inaccuracies	Data is reported with few inaccuracies	Data is accurately reported	
	Data analysis	I analyze some of the needed data	I analyze most of the needed data	I analyze all the needed data	
		Data analysis has frequent inaccuracies	Data analysis has few inaccuracies	Data analysis is accurate	
Conclusions	Conclusions	The conclusions I take are not consistent with the results	The conclusions I take are consistent with the results, but there are no references to the appropriate findings	The conclusions I take are consistent with the results and they are explicitly based on the findings	
	Limitations	I compare the findings with the expected results and suggest possible errors, not indicating their source	I compare the findings with the expected results, presenting errors and indicating some of their probable sources	I compare the findings with the expected results, presenting errors, indicating their probable sources and their effect in the results	I compare the findings with the expected results, presenting errors, indicating their probable sources and their effect in the results. I suggest alternative procedures to minimize them
	Learning outcomes	I describe incompletely what I have learned	I explain synthetically what I have learned	I explain what I have learned with detail	

Each section of the report (introduction, planning, procedure, results, conclusions) shall be analyzed concerning the organization, grammar, spelling and scientific style:

Category	1	2	3
Organization	I present texts, calculations, tables, diagrams and graphics in an unorganized way	I present texts, calculations, tables, diagrams and graphics in a reasonably organized way	I present texts, calculations, tables, diagrams and graphics in a very organized way
Grammar and spelling	There are frequent grammar and spelling mistakes	There are few grammar and spelling mistakes	Grammar and spelling are correct
Scientific style	Scientific terminology has some inaccuracies	Scientific terminology has few inaccuracies	Scientific terminology is accurate

It shall be done a holistic analysis of the report concerning the structure and the presentation.

Category	1	2	3
Structure and presentation	The report does not follow the proposed structure	The report follows the proposed structure, but different sections are not easily distinguished	The report follows the proposed structure and different sections are easily distinguished

References

Allal, L. (2007). Régulations des apprentissages: Orientations conceptuelles pour la recherché et la pratique en éduction. [Regulation of learning: Conceptual orientations for research and practices in education]. In L. Allal, & L. M. Lopez (Eds.), *Régulation des apprentissages en situation scolaire et en formation [Regulation of learning in school and training situation]* (pp. 7–23).Brussells: De Boeck & Larcier s.a..

Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. Assessment &. *Evaluation in Higher Education, 32*(2), 159–181. http://dx.doi.org/10.1080/02602930600801928.

Andrade, H., Du, Y., & Mycek, K. (2010). Rubric-referenced self-assessment and middle school students' writing. *Assessment in Education, 17*(2), 199–214. http://dx.doi.org/10.1080/09695941003696172.

Andrade, H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership, 57*(5), 13–18.

Andrade, H. (2001). The effects of instructional rubrics on learning to write. *Current Issues in Education 4*(4) . http://cie.ed.asu.edu/volume4/number4.

Bell, R., Smetana, L., & Binns, I. (2005). Simplifying inquiry instruction: Assessing the inquiry level of classroom activities. *Science Teacher, 72*(7), 30–33.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. http://dx.doi.org/10.1080/0969595980050102.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). *Working inside the black box: Assessment for learning in the classroom*, 1st ed. London: nferNelson.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. London: Open University Press.

Boekaerts, M., Maes, S., & Karoly, P. (2005). Self-regulation across domains of applied psychology: Is there an emerging consensus? *Applied Psychology*, 54(2), 149–154. http://dx.doi.org/10.1111/j.1464-0597.2005.00201.x.

Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).

Burns, R. B. (2000). *Introduction to research: Methods*, 4th ed. London: Sage Publications.

Clark, I. (2012). Formative assessment: A systematic and artistic process of instruction for supporting school and lifelong learning. *Canadian Journal of Education*, 35(2), 24–40.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education*, 5th ed. London and New York: Routledge/Falmer.

Crisp, G. (2012). Integrative assessment: Reframing assessment practice for current and future learning. *Assessment & Evaluation in Higher Education*, 37(1), 33–43. http://dx.doi.org/10.1080/02602938.2010.4942349.

Dann, R. (2002). *Promoting assessment as learning: Improving the learning process*, 1st ed. New York: Routledge Falmer.

Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education*, 24, 355–392. http://dx.doi.org/10.3102/0091732X024001355.

Green, R., & Bowser, M. (2006). Observations from the field: Sharing a literature review rubric. *Journal of Library Administration*, 45(1–2), 185–202. http://dx.doi.org/10.1300/J111v45n01_10.

Griffin, M. (2009). What is a rubric? *Assessment Update*, 21(6), 4–13.

Hadji, C. (1989). *L'evaluation, regles du jeu: Des intentions aux outils*. Paris: ESF.

Hafner, J., & Hafner, P. (2003). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education*, 25(12), 1509–1528. http://dx.doi.org/10.1080/0950069022000038268.

Hendry, G., Armstrong, S., & Bromberger, N. (2012). Implementing standards-based assessment effectively: incorporating discussion of exemplars into classroom teaching. Assessment &. *Evaluation in Higher Education*, 37(2), 149–161. http://dx.doi.org/10.1080/02602938.2010.515014.

Kirby, N., & Downs, C. (2007). Self-assessment and the disadvantaged student: Potential for encouraging self-regulated learning? *Assessment & Evaluation in Higher Education*, 32(4), 475–494. http://dx.doi.org/10.1080/02602930600896464.

Lichtman, M. (2006). *Qualitative research in education: A user's guide*. Thousand Oaks: Sage Publications.

Merriam, S. B. (1988). *Case study research in education: A qualitative approach*, 1st ed. S. Francisco: Jossey Bass.

National Research Council (1996). *National science education standards*. Washington, D.C: National Academy Press.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. http://dx.doi.org/10.1080/03075070600572090.

Nicol, D. (2010). From monologue to dialogue: improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, 35(5), 501–517. http://dx.doi.org/10.1080/02602931003786559.

Norton, L. (2004). Using assessment criteria as learning criteria: A case study in psychology. *Assessment & Evaluation in Higher Education*, 29(6), 687–702. http://dx.doi.org/10.1080/0260293042000227236.

Nunziati, G. (1990). Pour construire un dispositif d'évaluation formatrice. *Cahiers Pédagogiques*, 280, 47–64.

Papaleoutiou-Louca, E. (2003). The concept and instruction of metacognition. *Teacher Development*, 7(1), 9–30. http://dx.doi.org/10.1080/13664530300200184.

Price, M., Handley, K., Millar, J., & den Outer, B. (2007). *Engaging students with assessment feedback. Manuscript on case studies conducted for the FDTL5 project*. Oxford: Oxford Brookes University.

Rust, C., Price, M., & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and process. *Assessment & Evaluation in Higher Education*, 28(2), 147–164. http://dx.doi.org/10.1080/0260293032000045509.

Santos, L., & Gomes, A. (2006). Self-assessment and appropriation of assessment criteria. *Proceedings 30th psychology mathematics education (PME), mathematics in the center* (pp. 49–56)..

Santos, L., & Pinto, J. (2009). *Lights and shadows of feedback in mathematics learning. Thessaloniki, Greece, PME 33, International Group for the Psychology of Mathematics Education*, Vol. 5, 49–56.

Santos, L., & Semana, S. (2015). Developing mathematics written communication through expository writing supported by assessment strategies. *Educational Studies in Mathematics*, 88(1), 65–87. http://dx.doi.org/10.1007/s10649-014-9557-z.

Segers, M., Dochy, F., & Cascallar, E. (2003). The era of assessment engineering. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 1–12).Dordrecht: Kluwer Academic Publishers. http://dx.doi.org/10.1007/0-306-48125-1_1.

Semana, S., & Santos, L. (2010). *Self-assessment in written manuscripts. PME 34, International Group for the Psychology of Mathematics Education*, Vol. 4, Brasil: Belo Horizonte169–176.

Strauss, A., & Corbin, J. (1998). Basic of qualitative research. *Techniques and procedures for developing grounded theory*. Thousand Oaks, CA: Sage Publications.

Swaffield, S. (2011). Getting to the heart of authentic assessment for learning Assessment in Education: Principles. *Policy & Practice*, 18(4), 433–449. http://dx.doi.org/10.1080/0969594X.;1; 2011.582838.

Taras, M. (2001). The use of tutor feedback and student self-assessment in summative assessment tasks: Towards transparency for students and for tutors. *Assessment & Evaluation in Higher Education*, 26(6), 605–614. http://dx.doi.org/10.1080/02602930120093922.

Taras, M. (2003). To feedback or not to feedback in student self-assessment. *Assessment & Evaluation in Higher Education*, 28(5), 549–565. http://dx.doi.org/10.1080/02602930301678.

Tierney, R. (2013). Fairness in classroom assessment. In J. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 125–144).London: SAGE Publications, Inc.

Tillema, H. (2014). Student involvement in assessment of their learning. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 39–53).New York: Springer. http://dx.doi.org/10.1007/978-94-007-5902-2.

Vial, M. (2012). *Se repérer dans les modèles de l'évaluation: histoire, méthods, outils*. Bruxelles: De Boeck.

Vygotsky, L. S. (1934). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press [1934/1978].

Wharton, S. (2003). Defining appropriate criteria for the assessment of master's level TESOL assignments. *Assessment & Evaluation in Higher Education*, 28(6), 649–664. http://dx.doi.org/10.1080/0260293032000130261.

Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.

Woolf, H. (2004). Assessment criteria: Reflections on current practices. *Assessment & Evaluation in Higher Education*, 29(4), 479–493. http://dx.doi.org/10.1080/0260293031000168904_6.

Yin, R. (2002). *Case study research: Design and methods*, 3th ed. Newbury Park: Sage Publications.

Yore, L., Bisanz, G., & Hand, B. (2003). Examining the literacy component of science literacy: 25 years of language arts and science research. *International Journal of Science Education*, 25(6), 689–725. http://dx.doi.org/10.1080/0950069032000076661.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64–70. http://dx.doi.org/10.1207/s15430421tip4102_2.