Running head: Effects of co-creating rubrics

Co-creating rubrics: The effects on self-regulated learning, self-efficacy and

performance of establishing assessment criteria with students

Juan Fraile [1], Ernesto Panadero [2] & Rodrigo Pardo [3]

Author note

[1] Universidad Francisco de Vitoria, Madrid, Spain

[2] Departamento de Psicología Evolutiva y de la Educación. Universidad Autónoma de

Madrid, Madrid, Spain

[3] Departamento de Ciencias Sociales de la Actividad Física, del Deporte y del Ocio.

Universidad Politécnica de Madrid, Madrid, Spain

Correspondence concerning this article should be addressed to: Juan Fraile, Universidad

Francisco de Vitoria, Ctra. Pozuelo-Majadahonda Km. 1.800 – 28223 - Pozuelo de

Alarcón, Madrid - Spain.

E-mail: juan.fraile.@ufv.es

Abstract

The aim of this study was to compare the effects of co-creating rubrics against just using rubrics. By co-creating rubrics, the students might have the opportunity to better internalize them and have a voice in the assessment criteria. Two groups undertaking a degree in Sport Sciences (N = 65) participated. Results showed that the students who co-created the rubrics had higher levels of learning self-regulation measured through thinking aloud protocols, whereas the results from the self-reported self-regulation and self-efficacy questionnaires did not show significant differences. The treatment group outperformed the control group in only one out of the three tasks assessed. Regarding the perceptions about rubrics use, there were no significant differences except for the process of co-creation, to which the co-created rubric group gave higher importance. Therefore, this study has opened an interesting venue on rubrics research: co-creating rubrics may influence students' activation of learning strategies.

Keywords: rubric; self-assessment; self-regulation; self-efficacy.

1. Introduction

A rubric is usually defined as a document with a list of assessment criteria, a scoring strategy and quality definitions normally stated on a scale (Reddy & Andrade, 2010; Stiggins, 2001). Those standards definitions describe what students need to take into account to demonstrate a particular level of performance (Reddy & Andrade, 2010). Traditionally, rubrics have been used in a summative way as a tool for grading students' work (Andrade & Valtcheva, 2009; Jonsson & Svingby, 2007). However, recently rubrics have gained popularity because teachers provide them to students as a tool for formative assessment (FA), with the purpose of improving learning and performance. Panadero and Jonsson's (2013) review on formative rubric use finds that when rubrics are used with FA purposes the emphasis is on the communication of clear learning goals, success criteria, and provision of detailed feedback. A primary goal of formative rubric use is students' active use and internalization of the assessment criteria. It has been discussed that student involvement in rubric design/creation will facilitate their formative use of rubrics, rather than single-minded focus on the final score (Reddy & Andrade, 2010). Nonetheless, there is still a need for further empirical evidence to strengthen this claim. Therefore, this will be the aim of this study, exploring the effects of co-creating rubrics on students' performance, self-regulated learning, self-efficacy and perceptions about rubrics' use.

1.1. Formative use of rubrics

The use of rubrics can increase learning and performance under assessment for learning (AfL) and formative assessment conditions (e.g. enhancing self-regulated learning) (Panadero & Jonsson, 2013). These learning gains also come from aspects related to instructional purposes, such as teachers communicating their expectations for an assignment through the rubric, providing more detailed feedback and grading the

final product with higher reliability (Andrade, 2000; Moskal, 2003). On the other hand, if rubrics are used by teachers for summative purposes only (e.g. scoring the activity) the aim is no longer the students' learning, but there can still be an additional positive effect by enhancing the inter-rater and the rater (when only one teacher is scoring) reliability which results in more solid educational evaluation (Jonsson & Svingby, 2007).

Formative use of rubrics generally goes hand in hand with student self-assessment (Panadero & Jonsson, 2013), which denotes "the involvement of learners in making judgments about their own learning, particularly about their achievements and the outcomes of their learning" (Boud & Falchikov, 1989, p. 529). It is through self-assessment that students can reach a deeper understanding of their performance strengths and weaknesses, which allow them to improve over time (Kostons, van Gog, & Paas, 2012).

1.2. Rubric use for self-assessment and its effects on self-regulated learning and self-efficacy

Research has shown that there is a relationship between promoting self-assessment and students' self-regulated learning (SRL) (Kostons et al., 2012; Panadero & Jonsson, 2013), which is defined as "the sense of personal agency to enact this skill in relevant contexts. Self-regulation refers to self-generated thoughts, feelings, and actions that are planned and cyclically adapted to the attainment of personal goals" (Zimmerman, 2000, p. 14). Based on SRL models, there seems to be two SRL subprocesses linked directly to self-assessment: monitoring and self-evaluation (Zimmerman, 2000). Through the use of these two strategies students verify their progress and evaluate the outcome of the task. Scholars have further discussed that assessment criteria should be introduced during the SRL planning phase – before the

execution of the task starts – so that students can monitor and evaluate accordingly

(Andrade & Brookhart, 2016; Panadero & Alonso-Tapia, 2013). Sharing such criteria

can be achieved by introducing a rubric to students before their task execution.

However, it must be noted that while providing criteria (or a rubric for that matter) does

not guarantee their strategic use per se, self-regulation is more likely to occur when they

are provided (Lan, 1998).

It has been proposed that, for enhancing the strategic use of assessment criteria,

teachers should design activities that promote students' reflection about the learning

process (i.e. SRL). In other words, teachers should use self-assessment activities (Nicol

& Macfarlane-Dick, 2006). How can rubrics benefit such an aim? By using rubrics to

promote self-assessment, students will have access to the assessment criteria while they

are planning the task, which will lead to more realistic and adjusted learning goals

(Panadero & Alonso-Tapia, 2013). Then during their performance, they can monitor the

extent to which they are progressing in the desired direction using the rubric. Finally,

they will be able to self-evaluate their final product by using the rubric to reflect on how

they got there and what went right and wrong. All these processes should be modeled

by the teacher providing feedback on the self-assessment process itself (Andrade &

Valtcheva, 2009; Panadero, Jonsson, & Strijbos, 2016).

Another crucial learning variable is students' self-efficacy, which is the

confidence that students have in achieving a particular goal (Bandura, 2003). This

variable has been shown to be a strong predictor of academic performance (Richardson,

Abraham, & Bond, 2012), as students with higher levels of self-efficacy have higher

performance (Pajares, 2008). Furthermore, these students show more confidence,

intrinsic interest and perseverance in difficult tasks, leading to more efficient strategies

that improve learning while seeking the help of teachers and/or peers with a sharper

focus (Andrade, Wang, Du, & Akawi, 2009). The use of rubrics has been shown to increase students' self-efficacy (Andrade et al., 2009; Panadero, Alonso-Tapia, & Huertas, 2012; Panadero & Jonsson, 2013). This effect is probably based on handing out the rubrics to students beforehand, as when learning goals become clearer students have a better understanding of the learning target and how to achieve it. However, it has not yet been studied if co-creating rubrics would have an effect on self-efficacy over just using a rubric.

In summary, the process of students' formative rubric use, which involves goal-setting, planning, monitoring, and evaluating the final result, may improve SRL, self-efficacy and performance (Panadero & Jonsson, 2013). Then how can it be ensured that students actively use rubrics? A possible way may be involving then in rubric design and/or creation.

1.3. But, why co-create rubrics?

As previously stated, one of the keys to improving students' performance is that they must be aware of what is expected from them (Good, 1987), which can be achieved by formative uses of rubrics as pointed out above. However, as shown by Andrade and Du (2005) and replicated by Reynolds-Keefer (2010), students may perceive rubrics as instruments to reach the teachers' demands and standards. Therefore, rubrics can be perceived as external constraints to their learning with the only purpose of being giving the teachers what they want. Furthermore, one of the main criticisms of rubrics is that they can promote instrumentalism which leads to shallow approaches to learning (Torrance, 2007). This effect could be counteracted by involving students in the creation and negotiation of criteria which could improve their autonomy and empowerment. In fact, it has been argued that a better understanding of criteria and greater autonomy when applying such criteria can be reached by co-creating rubrics

(Andrade & Valtcheva, 2009; Panadero & Romero, 2014). In this regard, as long as students set their own goals and monitor their performance according to their criteria, they can self-regulate better in every context, therefore enhancing the possibilities to improve their academic achievement (Panadero, Brown, & Strijbos, 2016).

If students participate in the creation of rubrics, they are more likely to use this tool as if it belonged to their learning process. Otherwise, students could use rubrics just to know how scoring works (Reddy & Andrade, 2010), or because they represent what teachers want (Andrade & Du, 2005). This is supported by Kocakülah (2010) who noted that students could achieve a better grade as long as they were taught, and familiar with, rubrics. Thus, higher understanding and involvement can lead to increased motivation and confidence and therefore self-efficacy (Arter & McTighe, 2001).

Currently, only one study has explored the effect of co-creating rubrics on performance. In Kocakülah (2010), students assigned to the treatment condition created rubrics in groups of four, with each group creating a rubric. Students, under the supervision of the main researcher and rubrics experts, voted to select the best rubric. This rubric was then slightly modified and handed to the treatment students while the students in the control condition did not use a rubric. Results showed that the treatment condition outperformed the control. However, in this study it is impossible to disentangle the effects of the creating of rubrics and its use, as only the treatment groups used the rubric.

In a different line of work connected to the co-creation of rubrics, Andrade et al. (Andrade, Du, & Mycek, 2010; Andrade, Du, & Wang, 2008; Andrade et al., 2009) explored how discussing an exemplar affected performance. In these studies, the treatment conditions read a model essay (i.e. exemplar), discussed its strengths and weaknesses and listed quality aspects for effective writing. After that, a rubric

previously designed by the researchers was provided to students who self-assessed the first drafts with the rubric. The comparison group only listed qualities for an effective essay and reviewed their first drafts. These three studies reported greater performance in the treatment groups. Additionally, Andrade et al. (2009) also measured self-efficacy, through the Writing Self-Efficacy Scale, finding an increase for girls in the treatment group. Even though these three studies did not explore the co-creation of a rubric per se, they show evidence of the importance of discussing assessment criteria at the outset when performing a task.

In sum, the above-mentioned studies partially explored the effects of co-creation, finding a potential effect for learning and related variables (e.g. SRL, self-efficacy). However, these studies used a treatment group, which used rubrics, and a control group, which did not. This study aims to focus on the implications of a complete process of co-creation as the only difference between both groups because, here, the control group will also use the co-created rubric.

1.4. Aim, research questions and hypothesis

The aim of this study is to explore how co-creating rubrics (treatment group) compared to just handing out the same co-created rubrics (control group) might affect self-regulation, self-efficacy, performance and students' perceptions about rubrics.

Next we will present the logic that guides our research questions and hypotheses. Co-created rubrics have the aim of creating a deeper understanding of assessment criteria by students (Lim, 2013). This happens as a consequence of reflecting on and internalizing assessment criteria. Consequently, students set clear goals and activate proper learning strategies to self-regulate their learning (Kostons et al., 2012). Likewise, students increase the confidence in their own capacities as long as they have access to criteria (Andrade et al., 2009). Furthermore, Panadero and Jonsson

(2013) stated that the creation and use of co-created rubrics can enhance learning and performance.

Summing up, the research questions (RQ) and hypotheses (H) of this study are:

RQ1: Does co-creating rubrics enhance self-regulation? It is expected that the treatment group will outperform the control group (H1).

RQ2: Does co-creating rubrics enhance self-efficacy? The hypothesis is that the treatment group will benefit more showing higher self-efficacy (H2).

RQ3: Does co-creating rubrics enhance students' performance? It is expected that the treatment group will have higher academic performance (H3).

RQ4: Does co-creating rubrics affect the way rubrics are perceived? It is expected that students in the treatment group will be more critical, detecting weaknesses of this tool while they identify better its advantages and benefits (H4).

2. Method

2.1. Participants

The sample was comprised of 65 participants from two classroom groups: 34 in the co-created rubric condition (52.3%) and 31 in the rubric condition (47.7%) with the same teacher. Most participants were males (95.4%). The mean age was 23.4 years (SD = 2.48). The students were enrolled in a 'Sport in nature' course that belongs to the third year of the degree in Sport Sciences in a university in Spain. Participation in the study was voluntary but embedded in the instructional design of the course. Three actions were conducted to explore for differences among both conditions. First, the academic performance was controlled through the analysis of the participants' GPA. The analysis showed that there were no previous differences between both conditions [$F_{(1, 50)}$ =

.000, p = .99, η^2= .000; M_{co-cr}[1] = 6.71, M_{ru}[2] = 6.72]. Second, three questions explored

the participants' previous experience in the three tasks used in this study. No student

reported knowledge directly related to the three tasks. Third, previous rubric experience

was explored, with 11 students in the treatment group (32%) and 9 in the control group

(29%) reporting previous experience. Additionally, only one student in the treatment

group reported to have created a rubric before.

2.2. Materials

2.2.1. Instruments for assessing dependent variables

Self-regulated learning measures. In order to reach an appropriate estimation of

self-regulation, three different instruments were used following prior suggestions

(Boekaerts & Corno, 2005; Samuelstuen & Bråten, 2007):

Motivated Strategies for Learning Scales (MSLQ) (Pintrich, Smith, García, &

McKeachi, 1991). We used a tailored version of the MSLQ as significant results were

only expected in some of the scales. The initial version of the questionnaire that was

used to measure the conditions' SRL levels before the intervention included 37 five-

point Likert scale ('almost never'–'almost always') items. The following sub-scales

from the original questionnaire were used: intrinsic goal orientation (reliability index α

= .64), extrinsic goal orientation (α = .68), self-efficacy (α = .88), test anxiety (α = .67),

organization (α = .68) and metacognitive self-regulation (α = .76). The final version

used after the intervention included 51 items as we included scales (task value, control

beliefs about learning) in order that students could better reply after the intervention.

The combined reliability index was α = .77, with sub-scales ranging from .48 for

extrinsic goal orientation to .83 for self-efficacy.

---

[1] M_{co-cr}: Average of the group which co-created the rubrics (treatment group).

[2] M_{ru}: Average of the group which just handed out the same co-created rubrics (control group).

Specific Self-Regulation Questionnaire (SSR-Q). This questionnaire was created

for this study due to its specificity as the items needed to refer to the activity being

performed (Samuelstuen & Bråten, 2007). It is based on similar previous scales (e.g.

Panadero, Alonso-Tapia, & Huertas, 2014). It is composed of 15 five-point Likert scale

('almost never'–'almost always') items whose content refers to specific self-regulatory

actions related to the creation of sports activities. For example, 'Is this activity adequate

for the age of the participants?'. The reliability was $\alpha = .89$.

Thinking aloud protocols (TAP). This method implies that participants are asked

to perform a task and to verbalize what they are thinking about. It is considered an

appropriate representation of self-regulatory actions and metacognitive processes

(Ericsson & Simon, 1993; Greene, Robertson, & Croker Costa, 2011). Students were

video-recorded while self-assessing their own climbing activities. The content of the

videos was analyzed based on five categories:

- *General propositions:* the content did not have sufficiently specific information
  (e.g. 'Umm, that's funny!' referring to a climbing movement).

- *Rubric repetitive propositions:* literal repetition of a sentence from the rubric.

- *Rubric restated self-assessment propositions:* student self-assessed with their
  own words but just transforming the sentences of the rubric.

- *Self-regulatory propositions:* messages comparing their own climbing
  performance with the expert model, identification of successes and errors and
  their explanations.

- *Questions formulated:* number of questions students asked about the meaning of
  the quality definitions.

Self-efficacy questionnaire. It was created for this study to measure the students' activity specific self-efficacy, partly developed from previous research (e.g. Panadero et al., 2014). It includes 8 five-point Likert scale ('almost never'–'almost always') items about sports activities design (e.g. 'I think I am able to design activities which help participants to achieve the objectives'). Self-efficacy was measured before and after the intervention. The reliability index was $\alpha = .86$.

Performance measures. The study comprises three different measurement points for students' performance, all graded by the teacher using the co-created rubrics. First, an individual work about orienteering. The second and third topics, climbing and safety protocols, were graded through written tasks in the final exam of the course.

Students' perceptions about rubrics. A questionnaire (Appendix 1) was created for this study to measure students' perceptions about rubrics and their use. It included 13 five-point Likert scale ('completely disagree'–'completely agree') items with a reliability index of $\alpha = .62$.

2.2.2. *Instruments used for the intervention:*

Rubrics: Three analytic rubrics were co-created with the students for performing three different tasks (Appendix 2).

2.3. Design

This study has a quasi-experimental design with two experimental conditions (co-created rubric vs. control). Two of the dependent variables were only measured after the intervention: Performance (of three activities) and rubric perceptions. The other two dependent variables were measured pre and post: self-regulated learning and self-efficacy.

2.4. Procedure

The study took place in a one-semester course called 'Sport in nature'. This course belongs to the third year of a degree in Sport Sciences and two intact classroom groups participated. Both groups were taught by the same teacher, who was instructed to follow the same instructional structure and style. The first author presented this research to the participants in the first session indicating that the participation was voluntary and data would be treated confidentially. All the students accepted their participation in the study and filled out the questionnaires for self-regulation and self-efficacy.

Three analytic rubrics about the three different topics – orienteering, climbing and safety protocols – were co-created in different sessions by one classroom group chosen randomly. For co-creating the three rubrics, the following steps were followed:

1. After the topic was presented by the teacher, the students, in working groups of four, thought about the corresponding criteria for such an activity. Students listed criteria individually and then shared and discussed them within their group for another two minutes.

2. Every group communicated their list of criteria aloud and the teacher wrote them on the blackboard. Students, guided by the teacher, discussed in order to reduce the list to approximately eight final criteria.

3. Initially, three criteria were set to each group. Students in the same groups of four deliberated and wrote at least the two extreme quality definitions (a.k.a. 'poor' and 'excellent') for ten minutes. Each group had different criteria and two or three groups addressed each criterion separately.

4. Afterwards, the groups interchanged their members to share, discuss and combine the quality definitions drafted before. The group changes were made three times, three minutes each round. The teacher, who already had experience

with rubrics, helped the groups with the quality definitions until every group

agreed.

5. After class time, the teacher joined all students' contributions and created the co-
    created rubrics using the sentences and expressions of the students as much as
    possible. In the next session, teachers showed the resultant rubric to this
    classroom group in order to approve the final version all together.

6. Then, the teacher also gave the final co-created rubric to the non-co-created
    group with the same explanation.

The co-created rubrics were then handed out to the control group so that they
self-assessed with them too. This was done to ensure that the measured effect was the
co-creation, not the rubrics themselves. Consequently, both groups used the same
rubrics. The time spent creating the rubrics was around thirty minutes for each rubric.
The teacher employed this time in the non-co-created group to continue presenting the
three topics.

To measure students' self-regulation via thinking aloud protocols, students were
video-recorded performing the second topic – climbing – in the second session of this
topic. Six sessions later, students self-assessed watching their own videos while
thinking aloud.

Additionally, following Andrade and Valtcheva's (2009) recommendations, the
teacher implemented rubric-referenced self-assessment in the classroom which implies
that rubrics are used to enhance metacognitive activities in both groups for every topic
using the rubrics (i.e. students assessed climbing videos).

During the last session of the course, students filled out the self-regulation, self-
efficacy and rubric perceptions questionnaires. Lastly, they got the final exam of the
course in which students also used the rubrics.

3. Results

3.1. Effects of intervention on self-regulation

Self-regulation was measured using two questionnaires and thinking aloud protocols.

3.1.1. Motivated Strategies for Learning Scales (MSLQ)

The interaction intervention x occasion was not significant [F (1, 24) = .33, p = .572, η^2 = .01]. Both the co-created rubric group and the control group reported similar levels of self-regulation both before [F (1, 32) = .42, p = .52, η^2 = .013; $M_{co\text{-}cr}$ = 92.12, M_{ru} = 94.82] and after [F (1, 33) = .04, p = .85, η^2 = .001; $M_{co\text{-}cr}$ = 156.05, M_{ru} = 157.25] the intervention. In other words, the co-creation of rubrics did not have an impact on this measurement.

3.1.2. Specific Self-Regulation Questionnaire (SSR-Q)

No significant effects were found for the interaction treatment X occasion of measurement [F (1, 44) = .201, p = .66, η^2 = .005], neither for the occasion [F (1, 44) = .159, p = .69, η^2 = .004], nor for the intervention both before [F (1, 48) = .036, p = .85, η^2 = .001; $M_{co\text{-}cr}$ = 40.56, M_{ru} = 40.81] and after [F (1, 55) = .000, p = .99, η^2 = .000; $M_{co\text{-}cr}$ = 41.36, M_{ru} = 40.76].

3.1.3. Thinking aloud protocols

Regarding the five coding categories for the thinking aloud protocols, only two of them showed significant effects. The students who co-created the rubrics showed a higher level of self-regulation regarding 'self-regulatory propositions' than the control group [F (1, 18) = 6.93, p = .017, η^2 = .28; Dif. $M_{co\text{-}cr}$ = 3, M_{ru} = 1.95]. Besides, students who just received the rubrics – the control group – asked a significantly higher number of questions about the meaning of the quality definitions of the rubric [F (1, 18) = 8.86, p = .008, η^2 = .33; $M_{co\text{-}cr}$ = 0.25, M_{ru} = 1.25]. No significant differences were found in

general propositions [F (1, 18) = 3.84, p= .06, η^2= .176; M_{co-cr} = 1.83, M_{ru} = 3.25],

rubric repetitive propositions [F (1, 18) = .407, p= .53, η^2= .022; M_{co-cr} = 1, M_{ru} = 1.75],

nor rubric restated self-assessment propositions [F (1, 18) = 2.44, p= .14, η^2= .12; M_{co-cr}

= 5.25, M_{ru} = 3.5].

Summing up the results: out of the three SRL measurements, the self-reported

ones (MSLQ and SSR-Q) did not show significant results, while in the more objective

measurement, TAP, only two of the five coding categories favored the co-creating

group. Therefore H1 can be partially rejected.

3.2. Effects of intervention on self-efficacy

No significant effects were found for the interaction treatment X occasion of

measurement [F (1, 47) = 1.22, p= .28, η^2= .025]. However, even though the co-created

rubric group and the control group reported similar levels of self-efficacy before the

intervention [F (1, 50) = .44, p= .51, η^2= .009; M_{co-cr} = 25.54, M_{ru} = 24.86], after the

intervention the difference reached marginal significance [F (1, 55) = 3.57, p= .06, η^2=

.061; M_{co-cr} = 26.81, M_{ru} = 24.81]. These results can be seen in Figure 1. Moreover, the

occasion was significant for the co-created rubric group [F (1, 47) = 6.51, p= .01, η^2=

.122]. However, since the significant level is .06, H2 has to be rejected.

[Figure 1 near here]

3.3. Effects of intervention on performance

Students who co-created the rubrics only outperformed the control group in the

second task [F (1, 58) = 5.53, p= .02, η^2= .087; M_{co-cr}=6.44; M_{ru}=5.46]. No significant

differences were found regarding the first [F (1, 44) = 1.47, p= .23, η^2= .032; M_{co-cr}=7.04; M_{ru}=7.56] and the third tasks [F (1, 59) = 1.03, p= .31, η^2= .017; M_{co-cr}=6.42;

M_{ru}=5.95]. Thus, H3 should be partially rejected.

3.4. Effects of intervention on perceptions of rubrics

No significant effects were found except for the item related to co-creation of rubrics as can be seen in Table 1. Students who co-created the rubrics reported giving greater importance to the co-creation process than students of the control group [F (1, 55) = 39.79, $p < .001$, $\eta^2 = .42$; $M_{co\text{-}cr}$=3.65, M_{ru}=2.35]. Additionally, two other items had some significance, with the co-creating rubric students reporting a 'more objective and fair grade' and that the co-creation process was able to 'help them to self-assess'. Additionally, the co-created rubric group reached higher perceptions in all items except two. Therefore, the interpretation of H4 needs some reflection. In a strict interpretation, H4 has to be rejected. However, the use of co-created rubrics did enhance transparency for the students and help them in the process of self-assessment, even though two items did not reach significance ($p = .08$).

[Table 1 near here]

4. Discussion

The aim of this study was to compare the effects of co-creating and using rubrics, rather than just using the rubrics, on self-regulation, self-efficacy, performance and perceptions about rubrics. An intensive one-semester intervention was conducted based in the co-creation of three rubrics in the experimental condition and later use of three rubrics by the control and experimental conditions. It is important to point out that there is no previous research on such comparison, with only one study having explored the effects of co-creating rubrics against a control group that did not use any rubric (Kocakülah, 2010).

4.1. Self-regulated learning

It was hypothesized that the co-created rubric group would show higher use of self-regulated learning strategies because they would go through a deeper internalization of the assessment criteria and clearer goals for the task. This hypothesis

can be partially rejected as the three SRL measurements do not support it. First, MSLQ results showed no significant differences between conditions as both groups reported higher SRL levels after the intervention. Second, the specific SRL questionnaire (SSR-Q) did not show significant differences among the groups. Finally, the third measurement, thinking aloud protocols, showed significant differences benefiting the co-created rubric group in the two out of the five categories that referred to regulatory actions. Those two significant categories were the most relevant as they related to SRL propositions (i.e. SRL actions taken by the participants) and the number of doubts about the rubric. The result of this latter category implies that the group which did not co-create the rubric group required additional clarification regarding the quality definitions. Therefore, the co-created rubric group understood and better internalized the assessment criteria and standards from the rubrics. Thinking aloud protocols were used here to follow Boekaerts and Corno's recommendation (2005) about the use of situational measures for SRL. The reason is that thinking aloud protocols can measure more objective SRL aspects, where self-reporting cannot (Winne, 2010). Actually, the results of this study are in line with Panadero et al. (2012), in which significant differences were found only in thinking aloud protocols and not in self-regulation questionnaires. Thinking aloud protocols represent something external to the students' own measurements, whereas self-reported questionnaires assess students' self-regulation awareness and, therefore, it is largely affected by students' individual characteristics (i.e. cognitive load) (Panadero et al., 2012). In such a way, we can conclude that thinking aloud measures are more appropriate and specific for measuring self-regulation for the purpose of this study. Therefore, this study adds to the extended research on the effects of rubrics in self-regulated learning (see Panadero & Jonsson, 2013, for a review).

All in all, because the thinking aloud protocols are a more objective measure than self-reported ones (Greene et al., 2011) it could be argued that our hypothesis of co-creating rubrics increasing SRL could be maintained, as found in previous research with the use of rubrics (e.g. Panadero & Jonsson, 2013; Panadero & Romero, 2014). However, we have opted for a more conservative approach, partially rejecting the hypothesis and asking the reader to consider the above mentioned.

4.2. Self-efficacy

Our hypothesis that the co-creation of rubrics would enhance self-efficacy over the control group has to be rejected, however the difference between both conditions after the interventions was very close to significance (.06). This points out that, if the intervention had been longer or a higher number of participants were used, the difference might have been significant. Nevertheless, the effect is still considerably small. How do our results align with previous studies? On one hand, some studies did not find significant effects comparing a group of students which used rubrics and a control group (Panadero et al., 2014; Panadero, Alonso-Tapia, & Reche, 2013). On the other, the study of Panadero et al. (2012) showed that the students who used rubrics and received mastery feedback on three occasions increased self-efficacy. Moreover, Andrade et al. (2009) found an increase in self-efficacy for both long- and short-term interventions with larger effects from a long-term intervention on girls. Therefore, our results align with the two latest studies.

4.3. Performance

It was hypothesized that by co-creating the rubrics students would have reached higher performances. This hypothesis has to be partially rejected because the treatment group outperformed the control group in only one out of the three tasks assessed and, even in that one, the effect is extremely small. Nevertheless, one possible explanation is

that students benefited more from the thinking aloud activity as this was performed in the second task, which is the significant one, even if the effect was still small. By having an opportunity to reflect out loud, students in the co-created rubric group might have had more exposure to the use of their internalized criteria and standards.

4.4. Perceptions of rubrics

The hypothesis that the co-created rubric group would show more positive perceptions about rubrics use (H4) has to be rejected because the difference was only significant for 1 item out of 13. However, in the vast majority of items the co-created rubric group reported better non-significant perceptions about the use of rubrics. Particularly important is that two items directly related to transparency and understanding of standards (i.e. 'more objective and fair grade' and 'help to self-assess') were almost significant. In sum, it is probably the case that co-creating has an effect on students' perceptions about the use of rubrics, but a longer intervention might be needed for the effects to be deeper.

4.5. Limitations

There are four main limitations. First, the sample size is discrete. Second, thinking aloud protocols were only used in one of the three tasks. Therefore, there could have been a differential effect if that method had been used throughout the different tasks. Third, there was no control group without rubrics to also explore the effects of the co-created rubrics alone. However, our main aim was to explore the effects of co-creating rubrics, as there are just a few studies slightly related to co-creation but without control groups. And, fourth, there was a high percentage of males in the sample. Future research will need to explore the effects of co-creating rubrics with more gender balanced populations.

4.6. Future lines of research

First, more research related to the co-creation of rubrics is needed due to this being the first piece of research comparing this process to only the provision of the same rubric. Additionally, it is also necessary to explore the process of co-creation, seeking to maximize its virtues and reduce its limitations. Second, according to the results of this research, a longer intervention may produce significant effects in self-regulated learning, self-efficacy and performance. Third, more research is needed on self-regulated learning using thinking aloud protocols in order to obtain a more accurate measurement of their effect. Fourth, it would also be interesting to explore the effects on other variables such as accuracy or peer-assessment. Finally, the co-creation of rubrics and its effects should also be explored for topics other than physical tasks.

5. Conclusions

Co-creating rubrics could benefit self-regulation and performance rather than just using the same rubrics and, in longer interventions, could also enhance self-efficacy and students' perceptions about the use of rubrics. Importantly, this study employed a formative use of rubrics to enhance self-assessment as recommended in the literature (Andrade & Valtcheva, 2009; Panadero & Jonsson, 2013). Our results shed some light on rubric interventions where students have the opportunity to discuss the assessment criteria, standards and expectations, instead of having a rubric 'imposed' on them. Additionally, if rubrics are co-created, they can include language that is understood by the students which could exert a greater influence on the students using them and do so in a more strategic way. Therefore, this study has opened an interesting venue on rubrics research: the co-creation of rubrics, which needs more research to continue exploring and expand on our preliminary results.

References

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

Andrade, H., & Brookhart, S. M. (2016). The Role of Classroom Assessment in Supporting Self-Regulated Learning. In D. Laveault & L. Allal (Eds.), *Assessment for Learning: Meeting the Challenge of Implementation* (pp. 293–310). Springer International Publishing. http://doi.org/10.1007/978-3-319-39211-0

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

Andrade, H. L., Du, Y., & Wang, X. (2008). Putting Rubrics to the Test: The Effect of a Model, Criteria Generation, and Rubric- Referenced Self-Assessment on Elementary School Students' Writing. *Educational Measurement: Issues and Practice*, *27*(2), 3–13. http://doi.org/http://dx.doi.org/10.1111/j.1745-3992.2008.00118.x

Andrade, H. L., Wang, X., Du, Y., & Akawi, R. L. (2009). Rubric-Referenced Self-Assessment and Self-Efficacy for Writing. *The Journal of Educational Research*, *102*(4), 287–302. http://doi.org/10.3200/JOER.102.4.287-302

Andrade, H., & Valtcheva, A. (2009). Promoting Learning and Achievement Through Self-Assessment. *Theory Into Practice*, *48*(1), 12–19. http://doi.org/10.1080/00405840802577544

Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance*. Thousand Oaks, CA: Corwin/ Sage.

Bandura, A. (2003). *Self-efficacy: The exercise of control*. New York: W. H. Freeman.

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on

assessment and intervention. *Applied Psychology-an International Review-*

Psychologie Appliquee-Revue Internationale, *54*(2), 199–231.

Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in

higher- education: A critical analysis of findings. *Higher Education*, *18*(5), 529–

549.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*.

Cambridge, MA: MIT Press.

Good, T. L. (1987). Two decades of research on teacher expectations: Findings and

future directions. *Journal of Teacher Education*, *38*(4), 32–47. http://doi.org/http://

dx.doi.org/10.1177/002248718703800406.

Greene, J. A., Robertson, J., & Croker Costa, L. J. (2011). Assessing self-regulated

learning using thinking-aloud methods. In B. J. Zimmerman & D. H. Schunk

(Eds.), *Handbook of self-regulation of learning and performance* (pp. 313–328).

New York: Routledge.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and

educational consequences. *Educational Research Review*, *2*(2), 130–144.

http://doi.org/10.1016/j.edurev.2007.05.002

Kocakülah, M. S. (2010). Development and application of a rubric for evaluating

students' performance on Newton's laws of motion. *Journal of Science Education*

and Technology, *19*, 146–164. http://doi.org/10.1007/s10956-009-9188-9

Kostons, D., van Gog, T., & Paas, F. (2012). Training self-assessment and task-

selection skills: A cognitive approach to improving self-regulated learning.

Learning and Instruction, *22*(2), 121–132.

Lan, W. Y. (1998). Teaching self-monitoring skills in statistics. In D. H. Schunk & B. J.

Zimmerman (Eds.), *Self-regulated learning: From teaching to self-reflective practice* (pp. 86–105). New York, NY: Guilford Press.

Lim, J. M. A. (2013). Rubric-referenced oral production assessments : perceptions on the use and actual use of rubrics in oral production assessments of high school students of St . Scholastica´s College , Manila. *Language Testing in Asia*, *3*(4), 1–14.

Moskal, B. M. (2003). Recommendations for developing classroom performance assessments and scoring rubrics. *Practical Assessment, Research & Evaluation*, *8*(14). Retrieved from http://pareonline.net/getvn.asp?v=8&n=14

Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and selfregulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199–218.

Pajares, F. (2008). Motivational role of self-efficacy beliefs in self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning. Theory, research and applications* (pp. 111–168). New York: Lawrence Erlbaum Associates.

Panadero, E., & Alonso-Tapia, J. (2013). Self-assessment: Theoretical and Practical Connotations. When it Happens, How is it Acquired and what to do to Develop it in our Students. *Electronic Journal of Research in Educational Psychology*, *11*(2), 551–576.

Panadero, E., Alonso-Tapia, J., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences*, *22*(6), 806–813.

Panadero, E., Alonso-Tapia, J., & Huertas, J. A. (2014). Rubrics vs. self-assessment scripts: effects on first year university students' self-regulation and performance.

Infancia Y Aprendizaje, *37*(1), 149–183.

http://doi.org/10.1080/02103702.2014.881655

Panadero, E., Alonso-Tapia, J., & Reche, E. (2013). Rubrics vs. self-assessment scripts

effect on self-regulation, performance and self-efficacy in pre-service teachers.

Studies in Educational Evaluation, *39*(3), 125–132.

http://doi.org/10.1016/j.stueduc.2013.04.001

Panadero, E., Brown, G. T. L., & Strijbos, J. W. (2016). The Future of Student Self-

Assessment: a Review of Known Unknowns and Potential Directions. *Educational

Psychology Review*, *28*(4), 803–830. http://doi.org/10.1007/s10648-015-9350-2

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment

purposes revisited: A review. *Educational Research Review*, *9*, 129–144.

http://doi.org/10.1016/j.edurev.2013.01.002

Panadero, E., & Romero, M. (2014). To Rubric or Not to Rubric? The Effects of Self-

Assessment in Self-Regulation, Performance and Self-Efficacy. *Assessment in

Education: Principles, Policy & Practice*, *21*(2), 133–148.

Pintrich, P. R., Smith, D. A., García, T., & McKeachi, W. J. (1991). *A manual for the

use of the Motivated Strategies for Learning Questionnaire*. Ann Arbor, MI: The

Regents of the University of Michigan.

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education.

Assessment & Evaluation in Higher Education, *35*(4), 435–448.

http://doi.org/10.1080/02602930902862859

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of

university students' academic performance: A systematic review and meta-

analysis. *Psychological Bulletin*, *138*(2), 353–387.

http://doi.org/10.1037/a0026838

Samuelstuen, M. S., & Bråten, I. (2007). Examining the validity of self-reports on

   scales measuring students' strategic processing. *British Journal of Educational

   Psychology*, *77*(2), 351–378. http://doi.org/10.1348/000709906X106147

Stiggins, R. J. (2001). *Student-involved classroom assessment* (3er Ed.). Upper Saddle

   River, NJ: Prentice-Hall.

Torrance, H. (2007). Assessment as Learning? How the use of explicit learning

   objectives, assessment criteria and feedback in post-secondary education and

   training can come to dominate learning. *Assessment in Education: Principles,

   Policy & Practice*, *14*(3), 281–294.

Winne, P. H. (2010). Bootstrapping learner's self-regulated learning. *Psychological Test

   and Assessment Modeling*, *52*(4), 472–490.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In

   M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation*

   (pp. 13–39). San Diego, CA, US: Academic Press.

Figure 1. Interaction effect between group and occasion on self-efficacy.
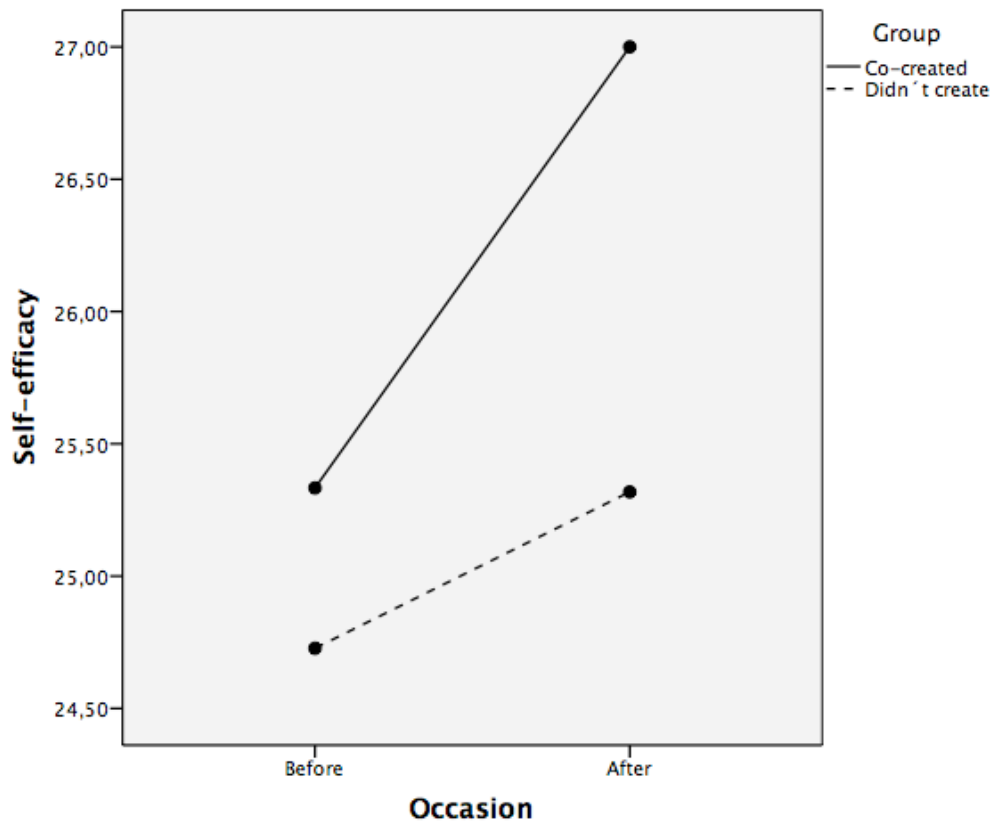
Appendix 1. English translation of questionnaire about students' perceptions

No.	Item
1	The use of rubrics helps me to understand the teacher's expectations
2	The use of rubrics helps me to plan and adjust my work according to the
3	learning goals
4	Using a rubric I can achieve a better grade than if I did not have it
5	The rubrics allow a more objective and fair grade on the part of the teacher
6	The use of rubrics helps me to optimize the time spent on my work and become more productive
7	The rubric helps me to self-assess
8	The use of rubrics helps me to create a work of a higher quality
9	When I use a rubric I only focus on the best quality definition for every criterion because that what is needed to get the best grade
10	The use of rubrics helps be not to become nervous in grading situations
11	Having a rubric makes me learn more than if I did not have it
12	I prefer a teacher who grades my work with a rubric, previously handed out to the students, rather than without using one.
13	In general, I think the use of rubrics is positive
	It is better to create the rubric with the teacher rather than the teacher hand it out directly

Appendix 2. English translation of the rubric used for the second topic: climbing.

		Poor 0%	Needs improvement 33%	Successful 66%	Excellent 100%
How to tie in	Figure 8 knot *2 points*	The knot does not ensure safety. It is not done properly and/or the end of the rope is too short.	The knot is tied incorrectly and/or the end of the rope is too long.	The knot is tied correctly and is approximately 15cm in length. However, it is not close enough to the harness.	The knot is tied correctly, with approximately 15cm in length and really close to the harness.
	Previous tasks *0.5 points*	Harness and/or helmet are not used properly or they are too loose.	Harness and/or helmet are not used properly. Harness is not over all clothes. A partner is not asked to check.	Harness is over all clothes and helmet is used properly, but a partner is not asked to check.	Harness is over all clothes and helmet is used properly. A partner is asked to check.
	To tie in *2 points*	The climber ties the rope to an incorrect point (i.e. rappel loop) of the harness. It is not checked.	The climber ties the rope correctly, but with any mistake (i.e. distance to the harness) while talking or being distracted. It is not checked.	The climber ties the rope correctly. However, it is not checked.	The climber ties the rope correctly and focused on it. Climber and belayer check everything.
Belay a lead climber	Until the first anchor *1 point*	The belayer is far from the climber and/or distracted.	The belayer is close to the climber but distracted, so he/she will not do his/her best to help the climber in case of a fall.	The belayer is close to the climber and focused, but he/she does not have an appropriate body position (i.e. arms up) to help the climber in case of a fall.	The belayer is close to the climber, focused and with an appropriate body position. Besides, the length of the rope is long enough to reach the first anchor.
	Rope *2 points*	The belayer does not manage the rope properly complicating the progress of the climber and increasing the risk of fall.	The belayer is relaxed and sometimes complicates the progress of the climber due to the short length of the rope.	The belayer is focused but does not have a proper control of the rope's tension and inappropriate climbing moves. If the climber should fall, the belayer could partially control the fall.	The belayer is focused, with both hands on the rope, controlling the tension of the rope and using proper moves. If the climber should fall, the belayer could control and cushion the fall.
	Body position *1.5 points*	The belayer is relaxed, distracted or sitting down. If the climber should fall, the belayer will not be able to control the fall.	The belayer is relaxed without paying attention most of time. If the climber should fall, the belayer would be dragged to the wall.	The belayer is not completely focused even if he/she has both hands on the rope. If the climber should fall, the belayer could partially control the fall.	The belayer has a proper body position and both hands on the rope. He/she uses the wall to balance if it is necessary. He/she will control the fall.
Belay a second climber	Belay a second climber *1 point*	The belayer is distracted and does not keep an appropriate tension on the rope increasing the risk of fall.	The belayer is distracted sometimes. He/she normally keeps an appropriate tension on the rope considering the situation and the requests of the climber.	The belayer keeps an appropriate tension of the rope considering the situation and the requests of the climber. However, he/she is not always focused, placing too much confidence in the belay device.	The belayer keeps an appropriate tension on the rope considering the situation and the requests of the climber. He/she is always focused, minimizing the risk of a fall.

Table 1. Rubric use perceptions questionnaire and statistical information.

# Item	$M_{co\text{-}cr}$	M_{ru}	Sig.	η^2
1. Understanding of teacher's expectations	3.45	3.15	.103	.048
2. Planning and adjustment of the work	3.33	3.31	.887	.000
3. Better grade	2.93	2.88	.810	.001
4. More objective and fair grade	3.53	3.23	.083	.055
5. Optimization the time spent	3.06	2.96	.582	.006
6. Help to self-assess	3.53	3.19	.080	.056
7. Higher quality work	3.20	3.28	.705	.003
8. Focus on the best quality definition	3.81	2.88	.418	.012
9. Not becoming nervous	2.68	2.54	.643	.004
10. Higher learning	2.94	2.77	.469	.01
11. Preference for grading with rubrics	3.26	2.88	.099	.049
12. The use of rubrics is positive	3.35	3.23	.528	.007
13. Preference for co-creating	3.65	2.35	.000	.42