Assessment in Education: Principles, Policy & Practice

ISSN: 0969-594X (Print) 1465-329X (Online) Journal homepage: http://www.tandfonline.com/loi/caie20

To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy

Ernesto Panadero & Margarida Romero

To cite this article: Ernesto Panadero & Margarida Romero (2014) To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy, Assessment in Education: Principles, Policy & Practice, 21:2, 133-148, DOI: 10.1080/0969594X.2013.877872

To link to this article: https://doi.org/10.1080/0969594X.2013.877872

Assessment in Education: Principles, Policy & Practice, 2014
Vol. 21, No. 2, 133–148, http://dx.doi.org/10.1080/0969594X.2013.877872

Routledge
Taylor & Francis Group

To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy

Ernesto Panadero[a,b]* and Margarida Romero[b]

[a]*Faculty of Education, Learning and Education Technology Research Unit, University of Oulu, Oulu, Finland;* [b]*Departament de Psicologia Bàsica, Evolutiva i de l'Educació, Universitat Autònoma de Barcelona, Barcelona, Spain*

(*Received 21 February 2013; accepted 8 December 2013*)

The objective of this study was to compare the effects of situations in which self-assessment was conducted using rubrics and situations in which no specific self-assessment tool was used. Two hundred and eighteen third-year pre-service teachers were assigned to either non-rubric or rubric self-assessment for designing a conceptual map. They then assessed their own maps. The dependent variables were self-regulation measured through a questionnaire and an open question on learning strategies use, performance based on an expert-assigned score, accuracy comparing self-scores with the expert's scores and task stress using one self-reported item. The results showed that the rubric group reported higher learning strategies use, performance and accuracy. However, the rubric group also reported more problems coping with stress and higher performance/avoidance self-regulation that was detrimental to learning.

Keywords: self-regulation; formative assessment; self-assessment; rubric; accuracy

A crucial aspect of becoming a successful and autonomous learner, especially as we move forward in the educational system, is the ability to monitor our own actions, thoughts and feelings to reach established goals. In order to do this, we must self-regulate our own learning (Winne & Nesbit, 2010; Zimmerman & Schunk, 2011). Among the different skills needed to become a self-regulated learner, various theorists have considered the strategies of self-evaluation and monitoring to be vital to success (Puustinen & Pulkkinen, 2001). Monitoring and self-evaluation show the student's capacity to judge their own performance and results, usually through close observation over the course of their performance and self-evaluation once the final result is reached (Zimmerman & Moylan, 2009). Not only are these strategies considered crucial for self-regulation, but empirical research has also found compelling evidence of their importance (Masui & De Corte, 2005). In a meta-analysis of self-regulation interventions, those that achieved the biggest effect made careful use of monitoring and evaluation in combination with planning (Dignath, Büttner, & Langfeldt, 2008).

Why are monitoring and self-evaluation so crucial? These skills are needed for students to be able to judge their own work. Without such reflection it may be

*Corresponding author. Email: ernesto.panadero@oulu.fi

difficult for them to distinguish between what they have done correctly and what they need to improve on (Kostons, van Gog, & Paas, 2012; Nückles, Hübner, & Renkl, 2009; Vermunt, 1998). The use of these two strategies by students is known as self-assessment, which involves internalising standards so that they can regulate their own learning, and are then able to evaluate their actions and have higher accuracy to detect their failures and how to correct them (Paris & Paris, 2001). Therefore, one of the key research questions for self-assessment research is: How can we help our students develop these two crucial skills?

Self-assessment

Self-assessment has been introduced in the classroom with the specific intention of promoting students' monitoring and self-evaluation strategies. Self-assessment has been defined as: 'the involvement of learners in making judgements about their own learning, particularly about their achievements and the outcomes of their learning' (Boud & Falchikov, 1989, p. 529). It has generally been accepted that learning to self-assess is a crucial step in implementing the kind of formative assessment which leads to a deeper understanding of the academic tasks and that students need to be taught how to self-assess (Andrade, 2010; Black & Wiliam, 1998; Dochy, Segers, & Sluijsmans, 1999; Kostons et al., 2012).

Heidi Andrade (formerly Goodrich) created a compelling list of conditions that need to be met in order for self-assessment to be effective:

> In order for effective self-assessment to occur, students need (according to Goodrich, 1996): awareness of the value of self-assessment, access to clear criteria on which to base the assessment, a specific task or performance to assess, models of self-assessment, direct instruction in and assistance with self-assessment, practice, cues regarding when it is appropriate to self-assess, and opportunities to revise and improve the task or performance. (Andrade & Valtcheva, 2009, p. 13)

An essential feature from the list of conditions that self-assessment instructional intervention should have is the provision of assessment criteria to students so that they are able to assess their work more objectively. For these criteria to have a bigger impact on the students' learning and strategies, they should be given to students before they start performing the task so that they can use the criteria to establish more realistic goals, monitor more accurately and evaluate the final product with more objectivity (Andrade, 2010; Panadero & Jonsson, 2013). Obviously, giving the students the assessment criteria does not guarantee that they will use them strategically, but it increases the chance that self-regulation would happen (Lan, 1998).

Approaches to self-assessment

Historically, there have been a number of different approaches to implementing self-assessment in the classroom (Dochy et al., 1999; Falchikov & Boud, 1989; Taras, 2010). Following the establishment of Andrade's criteria for self-assessment (Andrade & Valtcheva, 2009), Panadero, Tapia, and Huertas (2012) identified three different approaches: self-grading, the use of rubrics and the use of scripts/prompts/cues. In the present study, we will be focusing on the first two of these, as the third approach has been applied less often to real classroom situations (Panadero, 2011).

There is a range of studies that fall into the category of self-assessment through self-grading. These have progressed from the first attempts to influence students' learning by simply asking them to grade their work with no deeper reflection, which had a low learning impact (Boud & Falchikov, 1989), to more complex interventions that go beyond self-grading and request students to reflect deeply on their work (Dochy et al., 1999; McDonald & Boud, 2003; Taras, 2010). Even though the implementation of self-assessment has significantly improved over the years in this line of research, there remains one aspect that tends to be blurred: If students need assessment criteria (in the studies that mentioned them), how do they access them? In most of the previous research studies, this issue was not even discussed and when it was, the way that the criteria were accessed remained unclear.

With regard to the empirical evidence of the effects of this approach, Boud and Falchikov's review (1989) concluded that simply asking for self-grading did not guarantee that a learning effect would follow and that students were not generally accurate in terms of self-grading. As more recent studies (Landrum, 1999; Lejk & Wyvill, 2001) have also found, lower achievers tend to overestimate and higher achievers to underestimate themselves. Another review by Dochy et al. (1999) was carried out with a specific emphasis on the purpose of formative assessment. These authors concluded that students involved in self-assessment performed better in tests, learned more, were capable of a deeper understanding of their own work, achieved higher quality in their final tasks' products and took more responsibility for their own learning. Nevertheless, Tan's (2012) recent review on self-assessment research concluded that self-assessment produces mixed results that depend on the quality of the interventions and whether or not the conditions for adequate self-assessment are given. The latter refers to the presence of assessment criteria as described by Andrade and Valtcheva (2009).

The second approach to self-assessment, the use of rubrics, includes the assessment criteria as a distinct feature of these tools. Rubrics are documents that articulate the expectations of an assignment by listing the criteria for what is particularly important and by describing levels of quality on a scale from excellent to poor (Reddy & Andrade, 2010). Rubrics have three features: assessment criteria, a grading strategy and standards/quality definitions (Popham, 1997), which should help students reach acceptable performance standards by enabling them to establish adequate goals for tasks and grading these accordingly. One of the problems with rubric research is that it is not always used for formative assessment purposes, which clearly reduces its learning impact (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013).

A number of studies on the effects of rubrics on learning and performance have found a positive effect from their use when the intervention was adequate (e.g. applied for long enough) (Panadero & Jonsson, 2013) and accomplished most of the recommendations for self-assessment set out by Andrade (Andrade & Valtcheva, 2009; Goodrich, 1996), as merely handing out a rubric does not in itself guarantee a learning impact on students (Jonsson & Svingby, 2007). Rubrics have been found to have an impact on learning (Hafner & Hafner, 2003; Panadero et al., 2012) and performance (Andrade, Du, & Mycek, 2010; Andrade, Du, & Wang, 2008; Goodrich Andrade, 2001; McCormick, Dooley, Lindner, & Cummins, 2007; Panadero, 2011). Their use increases self-regulation compared to control groups and decreases negative self-regulation that is oriented to performance and avoidance goals (Panadero & Jonsson, 2013). Other important positive features of rubrics are their

capacity to increase self-assessed score accuracy (Jonsson & Svingby, 2007) and to increase self-efficacy on occasions where their use is combined with feedback (Andrade, Wang, Du, & Akawi, 2009; Panadero et al., 2012).

Another important aspect of rubrics is what are the students' perceptions about their use. All the studies exploring students' perceptions have found positive results, with rubric use decreasing anxiety for students and helping them feeling more secure among other effects (Andrade & Du, 2005; Reynolds-Keefer, 2010), and also being better liked when compared to another type of assessment tool script (Panadero, Alonso-Tapia, & Huertas, 2014). Nonetheless, sometimes students perceived that the rubric involved components that the teachers considered as important, but lacked other aspects that were relevant to them (e.g. students would like the rubric to include criteria regarding effort). To avoid this flaw, the rubric should be designed with the students who are the end-users (Andrade & Du, 2005). Students and teachers both benefit from the use of rubrics; the use of rubrics has a beneficial effect on teachers by helping them to clarify their assessment criteria and score fairly (Jonsson & Svingby, 2007; Schafer, Swanson, Bene, & Newberry, 2001).

However, rubrics also have additional flaws: students do not always value their use and find that they lack aspects that should be considered towards their grades (Andrade & Du, 2005; Reynolds-Keefer, 2010), and that the performance/learning positive effects are not always present (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013). These flaws can be avoided if students participate in rubric creation by negotiating for criteria that are relevant to them following Andrade's conditions (Andrade & Valtcheva, 2009), as simply handing rubrics out does not guarantee success (Jonsson & Svingby, 2007).

Thus, the self-assessment approaches of self-grading and rubric usage are the two most commonly used instructional techniques for promoting student monitoring and self-evaluation strategies. They present salient features that are clearly distinguishable and are based on the different effects that can be expected from learning. However, the comparison of their effects has remained unstudied and therefore it is needed to determine which of the two approaches produces the highest effect on student self-regulation and learning. For this reason, the aim of the present study is to compare the effects of self-assessment without giving the assessment criteria (self-grading) with the use of rubrics for self-regulation, performance, accuracy of self-score and task stress. This is done through training the students to do conceptual mapping, which is a learning strategy that increases student performance and is an effective technique to evaluate their knowledge (Berry & Chew, 2008; Jacobs-Lawson & Hershey, 2002).

Research questions and hypotheses

Here, we explore how the use of rubrics vs. simply asking students to self-assess might affect self-regulation, performance, accuracy and task stress. Firstly, it can be expected that using a rubric with assessment criteria that are clearly stated would affect the activation of self-regulatory and learning strategies. This would occur as the students would have a clearer understanding of what they have to achieve and how to reach that performance; previous research has found that this understanding affects students' self-regulation (Panadero & Jonsson, 2013). We will measure two types of self-regulation that have been previously explored in the rubrics literature (e.g. Panadero et al., 2012): learning self-regulation which refers to regulatory

actions oriented to learning goals (e.g. planning a task carefully before starting it) and performance/avoidance self-regulation which refers to actions guided by goals centred on performing or avoiding the task. These two types of goals are based on Boekaerts's top-down and bottom-up model (Boekaerts & Corno, 2005), with learning self-regulation being positive for learning goals, and performance/avoidance self-regulation being detrimental to learning (Alonso-Tapia, Panadero, & Ruiz, in press). We propose that students using rubrics will report higher levels of learning self-regulation after the intervention (H1) and lower levels of performance/avoidance self-regulation (H2). Students using rubrics will have clearer goals for the task, will be able to design a conceptual map using the rubric assessment criteria and therefore activate more learning strategies which will enhance their self-regulation use, as previous research has indicated (Kostons et al., 2012; Kramarski & Michalsky, 2010).

Secondly, as already mentioned, there is enough evidence to support the claim that the use of rubrics enhances learning and performance (Panadero & Jonsson, 2013). Nevertheless, the use of rubrics has never been compared with the effect of self-assessing without assessment tools, as we will explore in the present study. It is expected that the rubric condition will outperform the non-rubric condition in terms of performance (H3).

Thirdly, self-assessment accuracy usually refers to the deviation that occurs between the self-score given by the student and the one given by an evaluator (teacher, expert or peer). Typically, the accuracy of self-scores is reported in terms of under- or over-scoring when compared to the unbiased expert score (Wang & Imbrie, 2010). Self-assessment accuracy is important as it affects the validity of the assessment itself and the confidence that teachers and students have in this procedure (Brown, Glasswell, & Harland, 2004). It is then relevant to explore whether or not using rubrics would increase the accuracy in comparison to just asking the students to self-assess. It is expected that students using rubrics will be more accurate in self-grading their conceptual maps when compared with experts' scores (H4) due to the presence of the assessment criteria.

Lastly, students experience anxiety and stress when they have to self-assess, which negatively impact on learning (Andrade & Du, 2005). There is evidence that rubrics decrease these feelings (Panadero & Jonsson, 2013), and therefore it is relevant to compare their effect to those from self-assessment without tools. It is expected that students using rubrics will report lower levels of task stress (H5) as they will have a clear understanding of the standards and of their score, being less worried and anxious during the task.

Method

Participants

Two hundred and eighteen third-year pre-service teachers participated in this study. There were 189 women (86.7%) and 29 men (13.3%). The large presence of women is representative of the pre-service teacher programmes in Spain. The mean age was $M = 22.17$ (SD = 3.92). The students were enrolled in a 'Learning and Development II' course in a public university in Barcelona. They attended four different groups, each with a different teacher. Two groups were randomly selected for the non-rubric condition ($N = 107$) and the rest for the rubric condition ($N = 111$). The participants' previous experience on the task was controlled through evaluations of previous conceptual maps designed two months earlier in the same course.

Materials

Instruments for assessment

'Emotion and Motivation Self-Regulation Questionnaire' (EMSR-Q). (Alonso-Tapia et al., in press) (Appendix 1). This questionnaire includes 20 items that are answered on a five-point Likert scale, ranging from 'almost never' to 'almost always'. The items are grouped into two scales, *Learning self-regulation*, with eight items (α = .90), and *Performance/avoidance self-regulation*, with 12 items (α = .88). The first scale includes actions that are oriented to learning goals (e.g. 'I will plan the activity before starting to execute it'). The higher the value in this scale, the more positive the effect of self-regulation on learning would be. The second scale includes actions that show a lack of self-regulation or orientation to performance (e.g. 'I am getting nervous. I don't know how to do it'). The higher the value on this scale, the more negative the actions for learning performance would be.

Self-regulation open question. The participants were asked to report the strategies they used to design the conceptual map with no space limit, immediately after completing the self-regulation questionnaire. Three evaluators discussed and coded the answers of 60 participants to create the different categories for analysing this data (Krippendorff's α = .83). A total of 12 strategies were identified and organised around the three self-regulation phases, according to Zimmerman's model (Zimmerman & Moylan, 2009). The first phase was forethought (planning) and included the following strategies: underlining the text to read, deep-processing reading strategies (e.g. re-reading), identification of concepts from the text, making a list with the concepts to include in the conceptual map and designing a planning sheet. The second phase was performance and included the following strategies: making a draft, establishing relationships among concepts, visualising concept links, visualising a hierarchical structure and making a second, cleaner version of the conceptual map. Lastly, for the third self-reflection phase, two strategies were identified: re-reading the text after finalising the conceptual map and reviewing the conceptual map.

Task stress item. Task stress item. The purpose of this measure is to evaluate whether or not students experience different levels of stress while performing the task based on the rubric. Using a five-point Likert scale ranging from 'very low' to 'very high', the participants were asked about their levels of stress during the task.

Conceptual map. The first three independent evaluators scored 100 conceptual maps (Krippendorff's α = .86) using the same rubric as the participants in the experimental group – the control group did not have a rubric or any other tool with which to self-assess. The conceptual maps were chosen randomly by taking 50 maps out of each experimental group (rubric vs. non-rubric). Then one of the evaluators, the first author of this article who has ample experience evaluating conceptual maps, evaluated and scored the rest of the maps.

Instruments used for the intervention

Rubric. The rubric was created using expert models of conceptual maps (Appendix 2).

Text to read (task). The students read an excerpt from the text 'Estrategias docentes para un aprendizaje significativo' [Teachers' strategies for meaningful learning] by

Diaz Barriga (2002). The text is a mandatory reading from the official curriculum and describes the strategies that teachers can use to make their students' learning more meaningful and to activate deep learning approaches. The excerpt is 14 pages long and of an adequate level of difficulty for students in their third year of training – for example, some vocabulary is technical but well known by the students.

Design

A quasi-experimental approach was used with two types of designs depending on the measurement of the dependent variables. Firstly, as self-regulation was measured before and after the intervention, a 2 × 2 design was used: two experimental conditions – rubric vs. non-rubric – by two occasions – pre- and post-training. Secondly, factorial ANOVAs were used for the rest of the variables as they were only measured after the intervention, i.e. one between-group independent variable effect – rubric vs. non-rubric – for the following dependent variables: performance, accuracy and task stress.

Procedure

The participants were informed during the previous week that they were going to carry out a task during their usual session and that they needed to read a designated text. At the same time, they were given a URL link for filling in the self-regulation questionnaire online as a control measure.

During the session, the researcher gave the instructions for the task: to design a conceptual map based on the text given, which is an activity that would be assessed as part of their course grading. The rubric was handed out to the rubric group with an explanation on how to use it: first read the rubric, then perform the task while monitoring it with the rubric and finally look for each assessment criterion and grade their performance based on the different levels. The non-rubric group was told that they would have to self-assess once they finished their conceptual map. Then, the participants worked for 30 minutes to design their conceptual maps. After that, the researcher explained the procedure for self-assessment. The assessment criteria were read aloud to the non-rubric group: 'When assessing a conceptual map, an expert will consider the following features: all the relevant concepts have to be included, and the hierarchy has to be clearly defined …' and they were asked to score their own conceptual map. The rubric group was asked to score each rubric's criterion independently and then sum them up in a global score. Participants in both groups were encouraged to change their conceptual maps if they found room for improvement during their self-assessment. Finally, participants completed the self-regulation and self-efficacy questionnaires over the course of 20 minutes. The score and feedback were given to the participants in a later session.

Results

The participants' previous conceptual mapping experience was controlled by evaluating a conceptual map designed two months before the intervention in the same course. Three independent evaluators scored 40% of all the conceptual maps – one for each participant – using the rubric. The reliability of inter-rater agreement assessed through Krippendorff's α was .82. The rest of the conceptual maps were scored by one of the evaluators. The results showed that there were no significant

differences between the students' previous experience within both conditions ($p = .815$; Rubric $M = 12.64$, SD = 2.52; Non-rubric $M = 12.89$, SD = 2.42).

Effects of intervention on self-regulation

The results of the self-regulation questionnaire were mixed. With regard to the Learning self-regulation scale, the interaction INTERVENTION × OCCASION was not significant ($p = .139$; Rubric $M = 34.64$, SD = .564; Non-rubric $M = 32.54$, SD = .709), although the two main effects were significant. Therefore, the intervention effect was significant, with the rubric group reporting a higher level of Learning self-regulation ($F_{(1,150)} = 5.362$, $p < .05$, Rubric $M = 34.64$ [SD = .709], Non-rubric $M = 32.54$ [SD = .564]), and the occasion favouring later intervention ($F_{(1,150)} = 43.53$, $p < .001$, $\eta^2 = .225$, before $M = 35.78$ [SD = .493] and after $M = 31.40$ [SD = .622]), which shows that there was a higher level of self-regulation before the intervention than afterwards. Thus, the hypothesis H1 can be maintained.

With regard to the performance/avoidance scale, the interaction INTERVENTION × OCCASION was significant ($F_{(1,150)} = 7.62$, $p < .01$, $\eta^2 = .048$, Rubric $M = 26.87$ [SD = .684], Non-rubric $M = 24.59$ [SD = .859]). As can be seen in Figure 1, both groups experienced a decrease in this type of negative self-regulation, but the non-rubric group reported a lower level of this type of self-regulation after the intervention. Therefore, the hypothesis H2 has to be rejected.

The results from the open question about the types of strategies that the participants used favoured the rubric effect. Among all 12 strategies, there was a significantly higher reported use of the rubric condition with high effect sizes. The same tendency was reflected in the three categories for grouping the 12 strategies: forethought phase $F_{(1,216)} = 74.51$, $p < .001$, $\eta^2 = .257$, Rubric $M = 1.79$ (SD = 1.019), Non-rubric $M = .75$ (SD = .741); performance phase $F_{(1,216)} = 19.87$, $p < .001$, $\eta^2 = .084$, Rubric $M = 1.91$ (SD = .949), Non-rubric $M = 1.37$ (SD = .819); and self-reflection phase $F_{(1,216)} = 59.86$, $p < .001$, $\eta^2 = .217$, Rubric $M = .91$ (SD = .708), Non-rubric $M = .27$ (SD = .487). Therefore, the hypothesis H1 can still be supported in light of these results.

Intervention effects on performance

As hypothesised in H3, the rubric condition outperformed the non-rubric condition as measured by the students' scores on their conceptual maps given by experts:

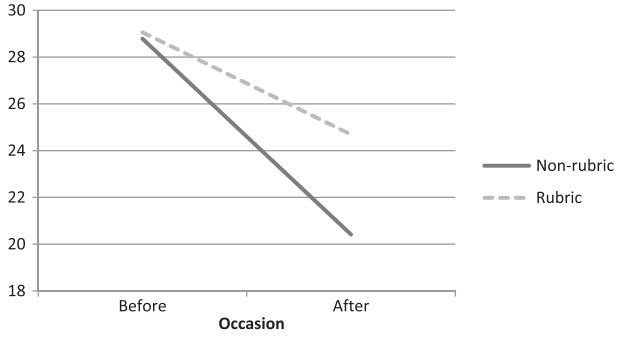

Figure 1. Interaction Intervention × Occasion effect on performance/avoidance self-regulation.

$F_{(1,210)} = 119.79$, $p < .000$, $\eta^2 = .363$, Rubric $M = 14.63$ (SD = 2.24), Non-rubric $M = 11.69$ (SD = 1.45).

Intervention effects on accuracy

The accuracy variable mean value represented the difference of the participants' self-assessment from the expert's grade. A lower value corresponded with greater accuracy of the participant's self-assessment. Both groups tended to overestimate their performance, but the rubric condition had significantly lower deviations which means that they were closer to the expert's grade compared to the non-rubric group: $F_{(1,209)} = 9.45$, $p < .001$, $\eta^2 = .043$, Rubric $M = .33$ (SD = 2.51), Non-rubric $M = 1.35$ (SD = 2.28). Therefore, the use of rubrics seems to favour self-assessment accuracy, as hypothesised in H4.

Intervention effect on perceived stress

The hypothesis for stress (H5) has to be rejected as the rubric condition reported a higher level of stress while performing the task: $F_{(1,216)} = 18.99$, $p < .000$, $\eta^2 = .081$, Rubric $M = 3.52$ (SD = 1.060), Non-rubric $M = 2.94$ (SD = .888).

Discussion

The aim of this study was to compare the effect of self-assessment without a rubric vs. self-assessment using a rubric for self-regulation, performance, self-efficacy, accuracy of self-grading and stress. It was hypothesised that the use of rubrics would enhance learning self-regulation, performance and accuracy while decreasing avoidance/performance self-regulation and stress. The majority of the hypotheses were maintained, which leads to the conclusion that rubrics are, when implemented well, valuable tools.

Self-regulation

Our results maintained that the use of rubrics enhances learning self-regulation more than simply asks students to self-assess. This conclusion comes from both the self-regulation questionnaire data and, in particular, the reported use of strategies for conceptual map design, with the latter showing a high effect size based on the intervention. Although both types of data were based on self-report from different tools, their validity is contrasting (Samuelstuen & Bråten, 2007): one adds to the other as they are not the same type of questions – one is Likert scale and the other open question. Therefore, it is probable that the use of rubrics has a considerable impact on self-regulation, as its use promotes the strategies that have been shown to have the biggest effect on self-regulation interventions: planning, monitoring and evaluation (Dignath & Büttner, 2008; Dignath et al., 2008). There is compelling research about the impact of self-assessment training on self-regulation, to which the present study added more evidence (Andrade, 2010; Kitsantas, Robert, & Doster, 2004; Kostons et al., 2012; Kramarski & Michalsky, 2010; Panadero et al., 2012). Most importantly, this study emphasised the relevance of adequate implementation of self-assessment in the classroom that fulfils Andrade's conditions (Andrade & Valtcheva, 2009; Goodrich, 1996). This study clarified that simply asking students

to self-assess their work is not as effective as using rubrics to enhance the use of positive self-regulatory strategies oriented to learning.

Nevertheless, the results from the performance/avoidance scale were not as encouraging regarding the use of rubrics, when based on a short intervention like the one used in the present study. According to our results, students using rubrics reported more actions that were detrimental to learning, as well as a higher level of stress while performing the tasks. These results were not in line with previous research in which the rubric condition always decreased performance/avoidance self-regulation to a higher extent than the other conditions (Panadero et al., 2014; Panadero, Alonso-Tapia, & Reche, 2013). However, in this study, the rubric group finished the intervention with a higher level of performance/avoidance self-regulation. What are the differences between these three studies? While the allocated time in the present study for the intervention was one hour, the learning task was carried out over an entire semester in the two previous studies. In shorter tasks, the participants could have experienced greater time pressure due to the tight deadline (Perlow, 1999).

In addition to time pressure, there is another explanation to these results. The participants knew that grades were important as their performance in the task would count towards their final score. To help diminish this negative effect, rubric interventions need to be longer so that the task itself would produce less time pressure. In this way, students can become familiar with the tool through practice and even receive guidance and example for using the rubrics. Regardless of this negative finding, the use of rubrics is still recommended due to the gains observed from the 'positive' self-regulatory strategy use, i.e. the strategies oriented to learning, and it is especially relevant for enhancing self-regulation among students in higher education, as previous research has shown (Ning & Downing, 2012).

Performance

Another positive effect from the use of rubrics was the improvement in student performance resulting from their use, which brings the present study in line with prior research (Andrade et al., 2008, 2010; Goodrich Andrade, 2001; Hafner & Hafner, 2003; McCormick et al., 2007). Another relevant finding was the high intervention effect size, which emphasises the crucial influence that rubrics can have on the performance of specific tasks (e.g. designing a conceptual map) and the benefits to students' learning that they can provide (Jonsson & Svingby, 2007). In summary, when rubrics are well designed, they can have a positive impact on performance because they set clear standards of how the final product of the task should look.

Accuracy

While some prior studies have found that students using rubrics can be accurate in the grades they give to their own performance (Hafner & Hafner, 2003; Jonsson & Svingby, 2007; McCormick et al., 2007; Sadler & Good, 2006) or their peers' performance (Panadero, Romero, & Strijbos, 2013), others have argued that they might not always be accurate (Falchikov & Boud, 1989; Tan, 2012). In the present study, we wanted to test the importance of knowledge of assessment criteria through a rubric on the accuracy of self-assessed scores. Our results were in line with previous research in that the use of rubrics has been shown to enhance the accuracy

of self-assessment and that it is also better to use a rubric than to ask for self-assessment without a tool, as the latter generally leads to inaccuracies. This effect is probably based on the rubric assessment criteria, i.e. the students share the same criteria as the teacher and thus know how to evaluate their work and improve accordingly (Andrade & Valtcheva, 2009). Thus, self-assessed scores can be reliable when a rubric – shared by the teacher and the students – is used, which gives evidence that the use of self-assessment grading through a rubric to reduce the burden of teachers assessing a high number of students can be very effective (Sadler & Good, 2006).

Limitations

One limitation of this study was the absence of a control group with which to compare results and to explore whether or not there was a gain, even for the non-rubric self-assessment condition. However, the focus of this study was the comparison between self-assessment with and without a rubric, as there has been compelling research contrasting rubrics groups vs. control groups. A second limitation was the assignment of participants to conditions that were only random at the group level. Even though we controlled for self-regulation and prior expertise in conceptual map design and found no significant differences, this limitation has to be kept in mind. Nevertheless, students were originally randomly assigned to their classroom groups by the university administration and there were a sufficiently high number of participants per condition. A third limitation of the present study was the use of self-reports to measure self-regulation, which depends more on the students' awareness of their own strategy use than other types of measures (Boekaerts & Corno, 2005). To counteract this limitation, quantitative and qualitative evaluations were done of both learning and performance/avoidance self-regulation measures, and these different evaluations were congruent and pointed to the same findings. Finally, a fourth limitation was the extraction of the sample from only higher education, so that cautions must be taken when transferring the significance of our results to other educational levels.

Educational and theoretical implications

Despite its limitations, our study resulted in several important findings for educational practice and showed that the use of rubrics can produce important gains in terms of self-regulation, performance and accuracy when compared to more traditional approaches of self-assessment (i.e. reading the assessment criteria aloud or simply asking students to self-assess). A recommendation that can be extracted from previous research would be to consider organisational characteristics and the implementation characteristics themselves to train self-assessment in classrooms (Schildkamp, Vanhoof, van Petegem, & Visscher, 2011). A recommendation of greater relevance from previous research would be to accompany rubrics with adequate conditions for self-assessment (Andrade & Valtcheva, 2009) in a way that avoids some of the detrimental effects of their use. In conclusion and in light of our results, the use of rubrics is highly recommended for students in higher education, especially if basic conditions are followed.

Funding

Research funded through grant to Ernesto Panadero by the Alianza 4 Universidades and EUROCAT project (FP7 Program).

Notes on contributors

Ernesto Panadero received his PhD degree in educational psychology by the Universidad Autónoma de Madrid (Spain) in 2011. He has worked as a postdoctoral researcher for one year at the Universitat Autònoma de Barcelona (Spain) and is currently working at the Department of Educational Sciences and Teacher Education within the Learning and Educational Technology Research Unit (LET) at the University of Oulu (Finland). His research covers aspects related to self-regulation and formative assessment.

Margarida Romero received the PhD degree in Psychology by UMR CNRS (France) and Universitat Autònoma de Barcelona (Extraordinary PhD Award in Psychology, Spain) in 2010. Since 2008, she has been the associate director of E-learning in ESADE Law & Business School and associate professor of psychology in UAB. From 2008 to 2012, she led the Euro-CAT-CSCL research project within the FP7 Marie Curie Industry-Academia Partnerships and Pathways (IAPP) actions. Her research aims to advance the understanding of the time factor in Computer Supported Collaborative Learning in the contexts of online learning and Serious Games.

References

Alonso-Tapia, J., Panadero, E., & Ruiz, M. A. (in press). Development and validity of the Emotion and Motivation Self-Regulation Questionnaire (EMSR-Q): Intertwined effects between self-regulation styles and classroom motivational climate. *Spanish Journal of Psychology, 17*.

Andrade, H. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In H. J. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 90–105). New York, NY: Routledge.

Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation, 10*(3), 1–11. Retrieved from http://pareonline.net/getvn.asp?v=10&n=3

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

Andrade, H., 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 Practices, 27*, 3–13.

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice, 48*, 12–19.

Andrade, H., Wang, X. L., Du, Y., & Akawi, R. L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *The Journal of Educational Research, 102*, 287–302.

Berry, J. W., & Chew, S. L. (2008). Improving learning through interventions of student-generated questions and concept maps. *Teaching of Psychology, 35*, 305–312.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*, 7–73.

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology, 54*, 199–231.

Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education, 18*, 529–549. doi:10.1007/BF00138746

Brown, G. T. L., Glasswell, K., & Harland, D. (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing Writing, 9*, 105–121. doi:10.1016/j.asw.2004.07.001

Diaz Barriga, F. (2002). *Estrategias docentes para un aprendizaje significativo: Una interpretación constructivista* [Teachers' strategies for a meaningful learning: A constructivist interpretation]. Mexico: McGraw-Hill.

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning, 3*, 231–264. doi:10.1007/s11409-008-9029-x

Dignath, C., Büttner, G., & Langfeldt, H. (2008). How can primary school students learn self-regulated learning strategies most effectively? A meta-analysis on self-regulation training programmes. *Educational Research Review, 3*, 101–129. doi:10.1016/j.edurev.2008.02.003

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education, 24*, 331–350.

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59*, 395–430.

Goodrich, H. W. (1996). *Student self-assessment: At the intersection of metacognition and authentic assessment* (Unpublished doctoral dissertation). Harvard Graduate School of Education, Cambridge, MA.

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

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*, 1509–1528. doi:10.1080/0950069022000038268

Jacobs-Lawson, J. M., & Hershey, D. A. (2002). Concept maps as an assessment tool in psychology courses. *Teaching of Psychology, 29*, 25–29. doi:10.1207/s15328023top2901_06.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*, 130–144.

Kitsantas, A., Robert, A. R., & Doster, J. (2004). Developing self-regulated learners: Goal setting, self-evaluation, and organizational signals during acquisition of procedural skills. *The Journal of Experimental Education, 72*, 269–287.

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*, 121–132. doi:10.1016/j.learninstruc.2011.08.004

Kramarski, B., & Michalsky, T. (2010). Preparing preservice teachers for self-regulated learning in the context of technological pedagogical content knowledge. *Learning and Instruction, 20*, 434–447. doi:10.1016/j.learninstruc.2009.05.003

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.

Landrum, R. E. (1999). Student expectations of grade inflation. *Journal of Research and Development in Education, 32*, 124–128.

Lejk, M., & Wyvill, M. (2001). The effect of the inclusion of self-assessment with peer assessment of contributions to a group project: A quantitative study of secret and agreed assessments. *Assessment & Evaluation In Higher Education, 26*, 551–561.

Masui, C., & De Corte, E. (2005). Learning to reflect and to attribute constructively as basic components of self-regulated learning. *British Journal of Educational Psychology, 75*, 351–372. doi:10.1348/000709905X25030

McCormick, M. J., Dooley, K. E., Lindner, J. R., & Cummins, R. L. (2007). Perceived growth versus actual growth in executive leadership competencies: An application of the stair-step behaviorally anchored evaluation approach. *Journal of Agricultural Education, 48*, 23–35.

McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education: Principles, Policy & Practice, 10*, 209–220.

Ning, H. K., & Downing, K. (2012). Influence of student learning experience on academic performance: The mediator and moderator effects of self-regulation and motivation. *British Educational Research Journal, 38*, 219–237. doi:10.1080/01411926.2010.538468

Nückles, M., Hübner, S., & Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. *Learning and Instruction, 19*, 259–271. doi:10.1016/j.learninstruc.2008.05.002

Panadero, E. (2011). *Instructional help for self-assessment and self-regulation: Evaluation of the efficacy of self-assessment scripts vs. rubrics* (Ph.D.). Universidad Autónoma de Madrid, Madrid.

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*, 806–813. doi:10.1016/j.lindif.2012.04.007

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).

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*, 125–132. doi:10.1016/j.stueduc.2013.04.001

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review, 9*, 129–144. doi:http://dx.doi.org/10.1016/j.edurev.2013.01.002

Panadero, E., Romero, M., & Strijbos, J. W. (2013). The impact of a rubric and friendship on construct validity of peer assessment, perceived fairness and comfort, and performance. *Studies in Educational Evaluation, 39*, 195–203. doi:10.1016/j.stueduc.2013.10.005

Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist, 36*, 89–101.

Perlow, L. A. (1999). The time famine: Toward a sociology of work time. *Administrative Science Quarterly, 44*, 57–81.

Popham, J. (1997). What's wrong and what's right with rubrics. *Educational Leadership, 55*, 72–75.

Puustinen, M., & Pulkkinen, L. (2001). Models of self-regulated learning: A review. *Scandinavian Journal of Educational Research, 45*, 269–286. doi:10.1080/00313830120074206

Reynolds-Keefer, L. (2010). Rubric-referenced assessment in teacher preparation: An opportunity to learn by using. *Practical Assessment Research & Evaluation, 15*(8). Retrieved from http://pareonline.net/getvn.asp?v=15&n=8

Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment, 11*(1), 1–31.

Samuelstuen, M. S., & Bråten, I. (2007). Examining the validity of self-reports on scales measuring students' strategic processing. *British Journal of Educational Psychology, 77*, 351–378. doi:10.1348/000709906x106147

Schafer, W. D., Swanson, G., Bene, N., & Newberry, G. (2001). Effects of teacher knowledge of rubrics on student achievement in four content areas. *Applied Measurement in Education, 14*, 151–170.

Schildkamp, K., Vanhoof, J., van Petegem, P., & Visscher, A. (2011). The use of school self-evaluation results in the Netherlands and Flanders. *British Educational Research Journal, 38*, 125–152. doi:10.1080/01411926.2010.528556

Tan, K. H. K. (2012). *Student self-assessment. Assessment, learning and empowerment*. Singapore: Research Publishing.

Taras, M. (2010). Student self-assessment: Processes and consequences. *Teaching in Higher Education, 15*, 199–209. doi:10.1080/13562511003620027

Vermunt, J. D. (1998). The regulation of constructive learning processes. *British Journal of Educational Psychology, 68*, 149–171. doi:10.1111/j.2044-8279.1998.tb01281.x

Wang, J., & Imbrie, P. K. (2010, June). *Students' peer evaluation calibration through the administration of vignettes*. Paper presented at the American Society for Engineering Education Annual Conference & Exposition, Louisville, KY.

Winne, P. H., & Nesbit, J. C. (2010). The psychology of academic achievement. *Annual Review of Psychology, 61*, 653–678.

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 299–315). New York, NY: Routledge.

Zimmerman, B. J., & Schunk, D. H. (2011). *Handbook of self-regulation of learning and performance*. New York, NY: Routledge.

Appendix 1. Items of the Emotion and Motivation Self-Regulation Questionnaire (EMSR-Q)

Source: Alonso-Tapia et al. (in press)

The student has to rate the frequency with which he/she experiences the thoughts and feelings listed below while performing conceptual maps, in a five-point Likert scale anchored from never to always.

Learning self-regulation, 8 items ($\alpha = .90$)

(1) This is going O.K.! … It seems that I understand it.
(2) Calm down … 'Do not hurry, do not stop' … You'll get it.
(3) Well … It seems that every time I do it better … I'm progressing …
(4) How interesting! It seems that I understand it.
(5) It is difficult, but how interesting! … I have to understand how to do it.
(6) This is not right … I'm going to check it step by step.
(7) How complicated! … Well, I'll continue … it is important to learn how to solve it.
(8) Here was the mistake! Great! Next time I know how to do it.

Performance/avoidance self-regulation, 12 items ($\alpha = .88$)

(1) This is not worth my time … Let's try to finish it as soon as possible.
(2) This task is a complete loss of time!
(3) What instructions so long! They only make me confused.
(4) What a boring task! Let's see if I finish and leave.
(5) I'm dead tired … Well, I had to continue to pass.
(6) I must go on … if I do not, I'll fail.
(7) What a mess! Well … Go on … if not you won't pass the exam.
(8) What a tiring task! … But I have to pass … Let's continue.
(9) What a stressful task! I'm doing it very bad … It's so difficult!
(10) This is so difficult … I am not going to be able to make it right.
(11) I am not made for doing this. If I could, I would give it up.
(12) I am getting nervous … I'm not able to do it.

Appendix 2. Rubric

Assessment criteria	Score			
	4	3	2	1
Concepts	All the important and secondary concepts are included	Contains the important and some secondary concepts but not all	The important concepts are included but not the secondary ones	Some key concepts are lacking
Hierarchy	The organisation is complete and correct and the map transmits it	The organisation is correct but incomplete: some levels or elements are lacking	The organisation is complete but incorrect: there are concepts in the wrong places	The organisation is incomplete and incorrect
Relationships among concepts in different hierarchical levels	RELATIONSHIPS They are correct making connections among the correct concepts LINKS Explicit and help to better understand the relationships among concepts	RELATIONSHIPS They are correct but incomplete: some connections are lacking LINKS Incomplete: Only some are explicit but they are correct	RELATIONSHIPS Some are incorrect making connections among concepts that do not have any relationship LINKS Only some are explicit but some are incorrect	RELATIONSHIPS The majority are incorrect or there are only a few LINKS Incomplete and incorrect
Relationships among concepts from different columns	There are several connections making relevant relationships	There is only one	None	None
Simplicity and easiness of understanding	Its design is simple and easily understandable	Some relationships are difficult to understand	There is an excessive number of connections	Neither the relationships nor the hierarchy are understandable
	There are examples	Contains a few examples	There are no examples	There are no examples