



Repositorio Institucional de la Universidad Autónoma de Madrid

<https://repositorio.uam.es>

Esta es la **versión de autor** del artículo publicado en:

This is an **author produced version** of a paper published in:

Studies in Educational Evaluation 39.3 (2013): 125–132

DOI: <http://dx.doi.org/10.1016/j.stueduc.2013.04.001>

Copyright: © 2013 Elsevier Ltd. All rights reserved

El acceso a la versión del editor puede requerir la suscripción del recurso

Access to the published version may require subscription

Studies in Educational Evaluation 39 (2013), 125–132

Rubrics vs. self-assessment scripts effect on self-regulation, performance and self-efficacy in pre-service teachers

Ernesto Panadero [a,b,*](#), Jesús Alonso-Tapia [a](#), Eloísa Reche [c](#)

a Departamento de Psicología Clínica y de la Salud, Universidad Autónoma de Madrid, Spain

b Learning and Educational Technology Research Unit, Faculty of Education, University of Oulu, Finland

c Departamento Pedagogía, Escuela Universitaria de Magisterio Sagrado Corazón, Córdoba, Spain

Correspondence concerning this article should be addressed to Ernesto Panadero, Departamento de Psicología Clínica y de la Salud, Universidad Autónoma de Madrid, Madrid, 28049, Spain.

E-mail: ernesto.panadero@gmail.com

Self-assessment scripts vs. rubric effect on self-regulation, performance and self-efficacy in university students.

Two approaches to self-assessment are optimal because they offer to the students the assessment criteria: rubrics and scripts. The aim of this study is to compare the effect of rubrics and scripts on self-regulation, performance and self-efficacy. 69 pre-service student teachers participated on the study. During a semester the participants were trained to design multimedia material. Results showed that students using script had higher levels of learning self-regulation when compare to the rubric and the control, and that the rubric group had higher level than the control. Regarding performance/avoidance self-regulation (negative self-regulatory actions detrimental for learning) the use of the rubric significantly decreased these negative actions. No significant effects were found for students' performance. A tendency was observed for students using self-assessment tools to have greater self-efficacy.

Keywords: self-regulation; self-assessment; formative assessment; rubric; script; self-efficacy; learning; goal orientation; self-grading; self-evaluation.

In recent years self-assessment of learning has received a lot of attention and it is a growing field in educational psychology (e.g. Dochy, Segers, & Sluijsmans, 1999; Ross, 2006; Taras, 2010). The reason is that self-assessment is a process necessary for self-regulation and learning to occur (Andrade & Valtcheva, 2009; Kitsantas & Peters, 2010; Winne & Hadwin, 1998). Consequently, researchers have looked for different ways to promote self-assessment in students. There are two instruments for promoting self-assessment, the potential effects and conditions for effectiveness of which are being studied: rubrics and scripts (Alonso-Tapia & Panadero, 2010). Rubrics are designed to evaluate, mainly, the product of an activity (Andrade, 2010; Jonsson & Svingby, 2007), whereas scripts are designed to help students during an activity to assess whether the process they are following is adequate or not (Bannert, 2009; Kramarski & Michalsky, 2010; Nückles, Hübner, & Renkl, 2009). These two tools have proved to have some positive effects on self-regulation and learning (Alonso-Tapia & Panadero, 2010; Bannert, 2009; Jonsson & Svingby, 2007). Nevertheless, results about script and rubric effectiveness and the conditions for it are far from being conclusive, and no prior research has compared their relative effects on self-regulation and learning in real classroom settings (Panadero, 2011). Hence, this will be the main objective of this study.

Theoretical framework

Our work is based on several theoretical suppositions about self-assessment and self-regulation. These processes – especially self-regulation – have received considerable attention in the last two decades and it is a crucial competence for higher education students to develop on the transition from secondary education (Torenbeek, Jansen, & Hofman, 2010) and for being successful during the university training (Heikkilä, & Lonka, 2006; Pintrich, 2004).

Self-regulation is a process through which self-generated thoughts, emotions and actions are planned and adapted to reach personal goals (Zimmerman, 2000). An important number of self-regulation theories point out that for such adaptation to occur persons must self-assess their on-going cognitive, emotional, motivational and behavioural processes. Doing this they can become aware of what needs to be controlled or changed (Kitsantas & Zimmerman, 2006; Winne & Hadwin, 1998; Zimmerman & Moylan, 2009). In their review of self-regulation theories, Puustinen and Pulkkinen (2001) point out that the five major self-regulation theories consider *self-assessment* a key self-regulation process, even though they refer to it using different names. There are also empirical findings that demonstrate the validity of these theories (Andrade & Valtcheva, 2009; Bannert, 2009; Heikkilä, & Lonka, 2006; Panadero, Alonso-Tapia & Huertas, 2012): when students self-assess their learning, problem solving, emotional and motivational processes using adequate criteria, they self-regulate their learning with success.

Conditions for adequate self-assessment.

However, what implies being able to self-assess one's own learning activity adequately? For students to be able to learn to self-assess there is a list of conditions for an adequate implementation (Andrade & Valtcheva, 2009). It can be extracted that for appropriate self-assessment to occur two factors crucial: (a) using adequate assessment criteria, and (b) using them at the opportune time (Panadero, 2011). Therefore, the questions to answer are: (a) what favours the use of adequate assessment criteria? and, (b) when is it opportune to use them?

Assessment criteria are the standards against which the execution process and final outcome of a task are evaluated. Though people can set their own assessment criteria for a task, students need to internalize the criteria provided by their teachers to carry out an

adequate self-assessment of their work. This internalization is difficult and, often, external help is necessary (Andrade, 2010).

As for the appropriate time, self-regulation is usually divided into different phases (e.g., Winne & Hadwin, 1998; Zimmerman & Moylan, 2009). According to the majority of theories (Puustinen & Pulkkinen, 2001), self-assessment takes part at the final phase – self-reflection phase (Zimmerman & Moylan, 2009) – where the students analyse what they have done and reflect about its consequences. Nevertheless, in line with other researchers (Boud, 1995; Winne & Hadwin, 1998), we consider that people cannot only self-assess the final product of the activity once it has finished, but also the process through which the final product is reached. In fact, according to self-assessment researchers (Andrade, 2010; Boud, 1995), a good implementation of self-assessment would influence the three self-regulatory phases: forethought, execution and self-reflection (Zimmerman & Moylan, 2009). This theoretical perspective is supported by research on the effects of self-regulation interventions showing that the biggest effects tend to occur when interventions focus on planning and monitoring or planning and evaluation (Dignath, Büttner, & Langfeldt, 2008). Therefore, the instructional help given should be used during the planning and monitoring phases of the self-regulation process and not only at the end of it.

Procedures for promoting self-assessment.

There are three types of interventions aimed at promoting self-assessment: (a) self-grading/self-evaluation or self-assessment without the assessment criteria, (b) rubrics, and (c) scripts, including cues and prompts (Alonso-Tapia & Panadero, 2010).

First, *self-evaluation* implies asking the student to evaluate their work and *score* it without using any specific tool. Research has shown that it is not a good pedagogical approach as it presents severe flaws (Dochy, Segers, & Sluijsmans, 1999; Falchikov & Boud,

1989). In this category can be also included those interventions aimed at enhancing self-assessment but that do not give the students the assessment criteria. As these two approaches do not include the assessment criteria they do not help to realize a precise self-assessment of task quality (Andrade and Valtcheva 2009).

Rubrics are self-assessment tools with three characteristics: a list of criteria for assessing the important goals of the task, a scale for grading the different levels of achievement and a description for each qualitative level (Andrade & Valtcheva, 2009). Rubrics have shown to enhance student performance and learning if used in combination with metacognitive activities (for a review: Panadero & Jonsson, 2013), to improve reliability among teachers when rating their students, and to improve reliability when the same teacher scores different students (for a review: Jonsson & Svingby, 2007), even though their direct effect on self-regulation needs more empirical evidence.

Scripts, including cues and prompts, are a specific set of statements or steps structured to follow the expert model of approaching a task from beginning to end. Like rubrics, they seem also to be promising tools to promote reflection and learning (Bannert, 2009; Peters & Kitsantas, 2010). They have been used mainly in experimental settings with only a small percentage of studies occurring in real settings (Kramarski & Michalsky, 2010; Panadero, 2011).

As rubrics and scripts contain the assessment criteria, they seem to be more effective methods than self-evaluation. However, what are the main differences between these two tools?

Differences between rubrics and scripts.

There are two main differences. First, rubrics have a scoring feature, therefore they put emphasis on grades whereas scripts do not have such characteristic. Second, rubrics usually

include a set of text-samples describing the characteristics that performance must have to achieve each grade, and thus they centre students' attention on outcomes or learning products, whereas scripts are formulated as questions pointing to the steps that the students have to follow, and thus they centre their attention on the learning process. In fact, when using scripts students have to use deeper metacognitive actions to understand whether they have carried out the correct action and reached the expected level. Therefore, both tools are oriented towards promoting students' self-assessment, but they present salient different features that can influence their effectiveness.

A comparison between rubrics and scripts was carried out by Panadero et al. (2012). They found that the participants using a script or a rubric had higher levels than the control group in self-regulation and learning, and that the use of the script enhanced self-regulation more than rubric. However, this study was conducted in an experimental setting with secondary education students, and in it learning was assessed using a task carried out just at the end of the intervention without effect for grades. Hence, it remained to test whether the intervention effects would be similar in natural classroom settings, with higher education students, and when learning was assessed using the tasks on which grades are based. That is the motivation of the present study.

Self-assessment training effects.

When planning the study, the new conditions raised an important question: Can different effects be expected from the use of each instrument in natural classroom settings?

First of all, it can be expected that *self-assessment help will positively affect learning self-regulation*. However, this effect may depend on the measurement method used. Measuring self-regulation is not an easy task. In a crucial article for the field, Boekaerts and Corno (2005) recommended using contextual measures of self-regulation instead of general

self-regulation questionnaires that may not be evaluating changes based on a specific intervention. This is in accordance with research that has proven that students do not always report accurately their use of self-regulation strategies (e.g. Winne & Jamieson-Noel, 2003). However, it is difficult to measure individual “on-line” self-regulation in natural classroom contexts unless students are working on a computer. Nevertheless, this difficulty can be at least partially overcome if self-regulation is measured through a combination of questionnaires: a general questionnaire assessing self-regulation messages and a specific one with items referring to the competence being acquired (Samuelstuen & Bråten, 2007). Regarding *self-regulation oriented to performance avoidance goals* (Boekaerts & Corno, 2005; Panadero et al., 2012) the self-assessment instruments should decrease these type of negative self-regulation in which the students regulate their actions to avoid the task or focusing just on the grade as the instruments would give them clear criteria de perform the task.

Second, if *self-assessment* affects self-regulation and learning in a positive way, *it may also produce an improvement in self-efficacy*, as some studies suggest (Alonso-Tapia & Panadero, 2010; Andrade, Wang, Du, & Akawi, 2009). Self-efficacy has an important role in two ways. First, pre-service teacher self-efficacy has been linked to their latter commitment during their adaptation to the workplace and performance (Klassen & Chiu, 2011). Second, more directly related to this study goals, self-efficacy has an essential role on self-regulation: if the students hold low expectations of self-efficacy for a specific task they are about to perform, their motivation will decrease and they will activate fewer and less effective self-regulatory processes; on the contrary, if they hold high self-efficacy expectations, they will be more willing to engage in highly demanding activities to overcome problems that they may find (Zimmerman & Moylan, 2009). If this were the case, the self-efficacy level prior to

intervention could mediate the effect of self-assessment in the final self-efficacy level. In fact, Alonso-Tapia and Panadero (2010) studied whether using self-assessment scripts enhanced self-efficacy in comparison to a control group based in previous research with rubrics that studied the same issue (Andrade et al., 2009). The latter had found that rubrics tended to enhance self-efficacy beliefs in interaction with gender (more so for girls than for boys). As there was no research available for the same effect in the case of scripts, Alonso-Tapia and Panadero (2010) studied this possibility, finding no significant effect. They attributed this result to the short duration of the intervention. Thus, the potential effects of scripts on self-efficacy remain an open question.

Third, though *graded performance* depends not only on self-assessment and self-regulation, it depends “also” on these factors. Therefore, *some improvement can be expected in performance as a result of improved self-assessment*.

Finally, to be motivated to use rubrics and scripts, students need to perceive their usefulness. Therefore, it is important to explore this perception and to analyse whether an increase in perceived usefulness correlates positively with self-regulation, self-efficacy and performance improvement.

The research *questions* and *hypothesis* of this study are thus as follows:

- (a) Do rubrics and scripts enhance self-regulation when compare to a control group? What self-assessment tool is more effective? Both tools will have positive effects on learning self-regulation over the control group (Hypothesis 1a). Scripts would enhance learning self-regulation in comparison to the rubric group because they promote deeper reflective processes (H1b). Regarding performance/avoidance self-regulation both self-assessment instrument groups would have lower levels of this type of negative self-regulation but no hypothesis about the comparison about rubric and scripts (H1c).

- (b) Do rubrics and scripts enhance performance over the control group? Both tools will have positive effects on performance compare to the control group as far as they have positive effects on self-regulation, though it may be that these are not high, according to the previous line of reasoning (H2).
- (c) Do rubrics and scripts enhance self-efficacy in comparison with the control group? If they do, what self-assessment tool is more effective? According to the previous line of reasoning there is no clear hypothesis (H3).
- (d) Which self-assessment tool, rubrics or scripts, do the students perceive as the better learning tool? According to the previous line of reasoning there is no clear hypothesis (H4).

Method

Participants

The sample was comprised of 69 participants: 20 in the rubric condition (29%), 20 in the script condition (29%) and 29 in the control group (42%). The majority of the participants were females: 58 females and 11 males scattered between the conditions (4 rubric, 3 script, 4 control). The mean age was: 20.6 years ($SD = 2.1$). Participation in the study was voluntary.

The participants were students in one of the three groups of the course “New technologies applied to education”. This course belongs to the second year of a teacher training program specializing in kindergarten/pre-school, a program with a high presence of female students. The main goal of this particular course is to prepare future teachers for the use of new technologies with pedagogical purposes. To accomplish this goal students have to learn how to design multimedia-learning material. These groups were selected because the students were going to acquire new skills and the intervention effects would have been less affected by the prior use of those skills.

The students had been randomly assigned to their natural classroom groups, although the researchers did not do this assignment. This fact could create differences in the characteristics of group composition. Accordingly, to assure group comparability, it was decided to control for goal orientation and previous experience in the design of multimedia material. Goal orientation has demonstrated being a predictor of self-regulation activation because depending on the goals the students are pursuing, they will be willing to activate the strategies needed to self-regulate when faced with difficulties (Pintrich, 2004). Due to its importance, goal orientation was measured to confirm whether all groups of students showed similar scores on goal orientations. In case they showed differences, goal orientation would be used as a covariate in the analyses.

Materials

Instruments for assessing dependent and moderating variables

Questionnaire of Learning Motivation and Expectancies (LEMEX) (Alonso-Tapia, Huertas, & Ruiz, 2010). This questionnaire was used for assessing goal orientations as moderating variables. It contains 178 items and measures goal orientations: learning ($\alpha = .88$), performance ($\alpha = .88$), and avoidance goals ($\alpha = .83$). The items used five-point Likert scales ranging from ‘completely disagree’ to ‘completely agree’.

Previous experience (PE). Two five-point Likert items were designed to measure previous experience in the design of multimedia material, as this variable could moderate the results.

Self-regulation measures. In order to reach a good estimation of self-regulation, following the advice of Boekaerts and Corno (2005), two different measures were used for assessing this process:

Self-regulation questionnaire (EMSR-Q) (Alonso-Tapia, Panadero & Ruiz, 2012). The

EMSR-Q is composed by 30 items organized around two second-order scales to be answered in a five-point Likert scale ranging from ‘never’ to ‘always’. The first scale, Learning Self-Regulation is formed by 13 items, and has reliability (Cronbach’s α) of .78. This scale measures self-regulating messages oriented to learning goals, for example: “I will plan the activity before starting to execute it”. The second scale, Performance/Avoidance Self-Regulation is formed by 17 items and has a reliability of .86. This scale measures self-regulating messages oriented to performance or the lack of self-regulation, for example: “I am getting nervous. I don’t know how to do it”. The higher the value in this scale the less positive the self-regulation for that student.

Specific self-regulation questionnaire (SSR). The SSR questionnaire was created for this study, and includes two five-point Likert scales ranging from ‘never’ to ‘always’. The first scale, seven items (originally nine items, two were discharged after the factor analysis), $\alpha = .81$ and explains 43.91% of the variance, includes questions and self-messages related to aspects specific to designing multimedia material using a Power-Point presentation as, for example: “Is this material I am designing easy to understand?”. The second scale, seven items (originally nine items, two were discharged after principal component factor analysis, as they loaded below .25, the standard level chosen for including or not an item in the scale), $\alpha = .81$ and explains 28.46% of the variance, includes questions and self-messages related to the specific actions that students have to complete for the course as the design of a WebQuest or Treasure Hunt –these are multimedia presentation, usually power points, containing learning tasks embedded in a game-. An example item is: “Are the steps to reach the end of the WebQuest/Treasure Hunt clearly defined?”. Both scales are answered on a five point Likert scale.

Self-efficacy questionnaire (SE). The SE was created for this study to analyze the

students' perceived self-efficacy towards the activities they were going to perform in the course. For that reason three different scales were created: (1) the scale *Didactic Unit Design* (six items, Cronbach $\alpha = .86$), that assesses the student's self-efficacy expectancies for designing a didactic unit according to learning and teaching principles; (2) the scale *Multimedia Material Design* (six items, $\alpha = .86$), that assesses self-efficacy for designing PowerPoint presentations; (3) the scale *WebQuest/Treasure Hunt* (six items, $\alpha = .89$), that assesses self-efficacy for designing either of these specific multimedia tasks. Pre and post measures were done for self-efficacy.

Performance. To measure performance, the Multimedia Material and WebQuest/Treasure Hunt designed by the participants were scored using the rubrics designed for this study. To assure inter-rater reliability two raters independently scored a sample of 20 students' activities. The level of agreement reached was 89%.

Perception of tool usefulness. Students in the rubric and script groups were asked how helpful rubrics and scripts were to perform the activities. This was a post measure.

Instruments used for the intervention

Self-assessment tools: Rubrics and scripts. Expert samples of PowerPoint presentations and WebQuest or Treasure Hunt were analyzed to create the rubrics and scripts that should help the students to self-assess their multimedia productions. After extracting the assessment criteria from these analyses, two rubrics and two scripts, one for each task, were developed. Rubrics were used by the teacher and one of the researchers to evaluate the students' final product for the performance measure (inter-judge agreement Kappa = .92).

Procedure

First, the research was presented to the participants in the first session of the semester pointing out that the participation was voluntary and that their data would be treated confidentially. No one declined to participate in the study.

Second, as the groups were natural classroom groups, it was not possible to assign the participants to the different conditions randomly. For this reason the participants' goal orientation was measured to check if the groups had participants pursuing similar goals. Therefore the participants completed the goal orientation questionnaire (LEMEX). A significant difference was found for performance goals as measured in the Performance Orientation scale and, in consequence, this data was used as a covariate in the latter analyses.

Third, students received instructions on the second session about how to design the multimedia material and the WebQuest/Treasure Hunt. Immediately afterwards, the self-assessment tools were handed out printed (Group A, rubrics; Group B, scripts; Group C, control) and the teacher explained how they should be used. After those instructions the students completed the general self-regulation and self-efficacy questionnaires. They also reported their previous experience designing PowerPoint presentations and WebQuest/Treasure Hunt.

The three groups had the same instructor who followed a specific planning that was done with the researchers to ensure that the three groups had the same pedagogical settings. This procedure was highly structured with clear and specific tasks for each lecture session, that were mainly based in a short introduction by the teacher about the tasks for the session and then individual work monitored by the teacher in the classroom. The teacher and one of the researchers had a meeting every two weeks to monitor the progress of the course development and discuss any possible deviation from the established standard procedure that there was not.

During the semester the teacher encouraged the use of the rubrics and scripts referring to these documents when giving feedback to the students in the different lectures. Also the students were reminded that the scripts and rubrics contained all the criteria that were needed to design the material. However, we did not try to gather information about the degree to which students used these aids because of the difficulty of controlling its accuracy.

Ten weeks/sessions later, the students submitted the multimedia material and the WebQuest/Treasure Hunt they had designed, along with the scripts and rubrics filled out by them including the students' self-assessment (e.g. their own scored in the rubrics categories). In the last session, they received their work scored, with feedback from the teacher. At that point they completed the general self-regulation, the specific self-regulation and the self-efficacy questionnaires. The rubric and script groups were also asked how helpful the self-assessment tool had been. Two weeks later the students took the exam compounded of twenty multiple choice questions (four options each) and four open questions. The exam represented 40% of the course grade with the other 60% divided equally between the PowerPoint score and the WebQuest/Treasure Hunt score. As mentioned, these were scored by the teacher and one researcher independently and then inter-judge agreement was calculated ($Kappa = .92$).

Analyses

First, all variables were screened for normality and linear relationships were calculated.

Second, it was tested if the three conditions were equal in their goal orientation and their previous experience designing multimedia material as these differences could have affected the participants' performance. A significant difference was found for performance orientation (see results) and therefore it was used as a co-variable to account for the difference between groups in the dependent variables in the rest of the analysis.

Third, in order to analyze the relative effects of each type of intervention, repeated measure ANCOVAs were used for the variables with pre and post measures (general self-regulation and self-efficacy). Factorial ANCOVAs were carried out for the variables measured only after the intervention (self-regulation, performance, perceived help of the tools and previous experience). Main effects and interactions were tested, and when necessary post-hoc Bonferroni tests were carried out.

Results

Preliminary analyses

Correlations between pre-intervention variables. These correlations are shown in Table 1. As can be seen, the majority of correlations are non-significant and those that reached significance had a low shared variance (less than 18%), except for r between the co-variables Learning Orientation and Avoidance Orientation, a result found also in previous studies (Alonso-Tapia, Huertas & Ruiz, 2010). Therefore the independent use of all these variables for the remaining analyses is supported.

Normality tests. All the normality tests were non-significant, though “Perceived Ability for WebQuest/Treasure Hunt” had a $p = .053$, a value that is close to reach significance. However, according to Ruiz-Maya (1977), the F statistic is robust enough to avoid being affected by small violations of normality, as in this case.

Covariates. A significant difference between conditions was found in Performance Orientation. Students from the control groups ($M = 49.43$) were significantly less oriented to performance than rubric ($M = 53.87$) and script ($M = 52.24$) students, $F(2, 58) = 4.78$; $p < .05$. Thus performance orientation was used as co-variable for the remaining analyses.

Previous experience in PowerPoint presentations and WebQuest/Treasure Hunt design. The three conditions did not differ on their levels of previous experience on the tasks they were

requested to perform for the course ($p = .65$). Therefore this variable was not used later as co-variable.

Intervention effects on self-regulation

Self-regulation was measured using two questionnaires in two occasions (pre and post): the EMSR-Q (comprised by two scales: Learning & Performance/Avoidance Self-regulation) and the specific self-regulation questionnaire (comprised by two scales: Multimedia Material and WebQuest/Treasure Hunt). The effect of the variable occasion was interpreted as an indicator of change

Learning self-regulation

The data for learning self-regulation (higher values in this variable mean that the students used more strategies oriented to learning goals) is coming from the learning self-regulation scale from the EMSR-Q and the two scales from the specific self-regulation questionnaire.

Data from the EMSR-Q learning self-regulation scale showed that the interaction effect between the occasion (pre-post) and the training (control vs. rubric vs. script) was significant, $F(2, 64) = 5.37$; $p < .01$, $\eta^2 = .168$. Differences are illustrated in Figure 1, post-hoc Bonferroni analysis showed that the difference between script and rubric before the intervention was significant (Dif.: 5.34; $p < .05$), but not after. This result was due to the fact that, after intervention, learning self-regulation increased in the script group significantly (Dif.: 3.01; $p < .06$), whereas in the rubric group decreased significantly (Dif.: -4.40; $p < .01$). The control group pre and post scores did not differ. Thus, according to our expectations, scripts have a positive effect enhancing learning self-regulation (H1b), but contrary to our expectations, there were no differences between the self-assessment conditions and the control group (H1a). An unexpected result was the decrease on learning self-regulation in the

rubric group.

Insert Figure 1.

Regarding the data from the specific self-regulation questionnaire there were no significant effects either on the Multimedia Material scale, $F(2, 69) = 1.451$; $p = .243$; $\eta^2 = .051$; Rubric Adjusted $M = 20.87$, Script Ad. $M = 20.93$, Control Ad. $M = 18.43$, or on the WebQuest/Treasure Hunt scale, $F(2, 69) = .866$; $p = .426$; $\eta^2 = .031$; Rubric Adjusted $M = 24.93$, Script Ad. $M = 24.27$, Control Ad. $M = 22.04$. Therefore, results run against the hypothesis that script and rubric would promote more learning self-regulation than the control (H1a).

Performance/Avoidance self-regulation

Data for this type of detrimental for learning self-regulation was coming from the EMSR-Q performance/avoidance scale. The interaction between occasion and training was significant, $F(2, 66) = 5.72$; $p < .001$; $\eta^2 = .175$. Differences are illustrated in Figure 2, post-hoc Bonferroni analysis showed that this effect is only due to the fact that the rubric-group score decreased in a significant way ($p < .001$) in their performance/avoidance self-regulation whereas the other groups remain with the same level. In the pre measure the difference between rubric and control groups was significant ($p < .05$). Therefore our hypothesis (H1c) can only be maintained partially as rubrics had a positive effect but this effect is null in the case of scripts.

Insert Figure 2.

Performance

No significant differences were found in performance either for Multimedia Material ($p = .063$) or for WebQuest/Treasure Hunt ($p = .864$), a result that runs against our expectations (H2).

Intervention effects on self-efficacy

Self-efficacy was measured before and after the intervention through three specific scales. It was found between groups differences in self-efficacy for designing Multimedia Material $F(1, 53) = 3.871$; $p < .05$, $\eta^2 = .127$; Rubric Adjusted $M = 19.16$, Script Ad. $M = 20.33$, Control Ad. $M = 17.4$. Post-hoc Bonferroni analysis showed that this effect is due to the difference between script and control groups (Dif.: 2.36; $p = .04$). Data from the other two scales –Didactic Unit Design and WebQuest/Treasure Hunt- did not show differences among the groups. It was not hypothesized the direction of this results (H3).

Perception of usefulness of self-assessment tool

After the intervention both self-assessment groups were asked to report their preference for their tool. The students using rubrics perceived their tool as more helpful than the students using scripts did, $F(1, 28) = 22.76$; $p < .001$; $\eta^2 = .477$; Rubric Adjusted $M = 4.6$, Script Ad. $M = 3.8$. It was not hypothesized the direction of this result (H4).

Discussion

The aim of this study was to test the effects of different self-assessment instruments (rubric and script) in self-regulation, performance and self-efficacy with university students in a natural context. What have been its contributions in relation to these objectives?

Self-regulation

This study results are interesting because they inform about aspects that should be clarified by future research and that can be used to determine different uses of rubrics and scripts. While scripts enhanced learning self-regulation, rubrics decreased performance/avoidance self-regulation pointing out that these instruments have different effects on the way the students are affected by their use. Next, we will explain these in more detail.

Regarding learning self-regulation, this study *partially* supports the notion that the use of scripts increases the self-regulatory messages oriented to learning. This was hypothesized based on previous research (Bannert, 2009; Kramarski & Michalsky, 2010; Peters & Kitsantas, 2010) because the purpose of scripts is to increase the metacognitive awareness and therefore to activate more learning strategies. We wrote “*partially*” because only the results from the EMSR-Q support this idea but not the data coming from the specific self-regulation questionnaire. However, the latest questionnaire can be affected by its content. It referred to regulatory actions specific for the task and, as they are needed to complete it, students from the three conditions might have performed them regardless of the presence of the self-assessment instruments. Thus, it could be difficult to draw any general conclusion about the effect on learning self-regulation. Nevertheless, our results should be interpreted in the context of results of similar studies. In a previous study, Panadero et al. (2012) used thinking aloud protocols to measure learning self-regulation. They found that scripts had the highest positive effect on learning self-regulation followed by the rubrics in comparison to the control group. In conclusion, the partial results found here and the clearer ones coming from a very similar research using a more objective measurement of self-regulation point out that scripts have a positive effect on learning self-regulation.

As for performance/avoidance self-regulatory messages, the results of this study support the notion that only rubrics contribute to their decrease. There is only one previous study that has explored this type of self-regulation and how it is influenced by promoting self-assessment. Panadero et al. (2012) used also a similar general scale with secondary students and found no significant differences based on the self-assessment tools effect. However, there is one crucial difference that might have affected the levels of performance/avoidance self-regulation differently in both studies, but not the levels of learning self-regulation. While in

Panadero et al. (2012) the study was conducted in an experimental setting and the students' performance was not related to a course grade, in the present study the performance of the students was critical for their final grade. Therefore, the participants in this study could have been more stressed and pressured to perform well because they had a grade pending on their work. For that reason, rubrics might have made a difference as they explicate how the students' performance will correspond with the later grade from the instructor. On the other hand, scripts are more cognitively demanding -the students need to activate more strategies to use the scripts properly-, but do not inform about how the performance relates to the final score. In sum, rubrics might be a better tool in tasks that will be graded, but this hypothesis should be tested by future research.

In conclusion, scripts seem to have the potential to activate the students' use of learning strategies while rubrics seem to have the potential to decrease the use of emotional self-regulation strategies that can have a negative effect on learning. One possible implication of these results can be that for activities in which the students might be experiencing anxiety (e.g. activities that have a deadline and will be counting for the grade) the rubric could help them to focus more on the learning process, while the script might have an advantage if used in complex activities that require deep processing guiding the students to activate more learning strategies.

Effects on Performance

Two related effects can explain the lack of effects from the self-assessment instruments on performance. First, performance was measured through the final score of the course and it might have been affected by factors beyond the use of self-assessment tools that were not controlled even if we equaled the pedagogical environments in the three experimental conditions –same teacher, same instructions, same activities agenda, etc.-.

Second, the structure of the activities was highly defined, and the control group may have benefited from these highly defined tasks. Therefore, the effects of the self-assessment tools could also have been diluted by these two non-controlled variables.

If we consider previous research, it is well documented that rubrics (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013) and scripts/prompts (Kramarski & Michalsky, 2010; Peters & Kitsantas, 2010) have positive effects in performance and learning. Therefore the lack of effects in this study can be explained by the type and nature of the task chosen, as it will be explained in more detail later.

Effects on Self-efficacy

There were no significant differences in self-efficacy change. An explanation would be that when students perform a task and obtain a final product, they might experience that “they are able” to do it. Therefore, it seems plausible that all students in our study maintained their perception of efficacy as all of them reached final products (a didactic Power Point and a WebQuest/Treasure hunt).

Then, under what conditions could self-efficacy be enhanced? There can be two explanations. First, it may be that self-efficacy increase depends on feedback, its frequency and characteristics. In the study by Panadero et al. (2012), students in the different conditions received feedback in three occasions, and all groups increased self-efficacy but more if feedback was on process rather than on performance. Besides, a significant interaction was found between type of feedback and self-assessment tool: the group that received rubrics and mastery feedback increased their efficacy perception more than the other groups. Therefore, the existence and type of feedback may explain changes in self-efficacy. In the present study, the scarcity of feedback – only once – and the nature of it – a score – is coherent with the explanation proposed.

Second, according to the review by van Dinther, Dochy and Segers (2010), for self-efficacy to improve, it seems necessary to provide students with practical experiences -i.e. students should perform a task while applying knowledge and skills within demanding situations-. Consequently, it may be that only if a task is carried out in such conditions – no matter whether rubrics or scripts are used –self-efficacy will improve. This second explanation is not incompatible with the first. In fact, when students apply their knowledge to a practical task, they can receive feedback on the quality of performance, and this feedback can influence their perception and expectancies of self-efficacy. Future work should address the effect of these two tools on self-efficacy using a stronger measure than the one used here.

Relationship between self-regulation, self-efficacy and performance results: the importance of the task

Rubric and script had a positive effect on self-regulation, but not on performance and self-efficacy. This fact rises the question of the relations between these variables. It seems that, although there is a tendency for rubrics and scripts to improve self-regulation in different ways, the relative magnitude of such effects and how this affects other variables (self-efficacy and performance) may be manifested or not depending on the type of task, measure and context. Higher levels of conscious self-regulation do not necessarily conduct to better performance if the task can be perform more efficiently with less reflection because not all tasks need the same deep metacognitive activity to produce positive outcomes, as they can be more or less automatized (Brown, 1987; Panadero et al., 2012; Reitmeier & Vrchota, 2009). The fact that in this study self-regulation was affected differently depending on the use of rubrics and scripts, but that self-efficacy and performance were not affected, points to this hypothesis. The tasks performed in this study were highly-structured, with very specific steps outlined, and the participants had previous experience performing them. Therefore, the lack of

effects on performance may have been due to task choice and to how the pedagogical setting was planned. Consequently, it is necessary that future research conduct a systematic study of the effects of these variables taking into account task and instructional context properties.

Tool perception

The fact that the rubrics were perceived as more helpful than scripts can have two explanations, one not excluding the other. First, rubrics could be promoting a feeling of security based on the scoring feature, as explained before. As it is known, higher education students are aware of the importance of grades (Pintrich, 2004). Thus, once the students finished the activities, those using rubrics could have some certainty about their grade based on the use of the scoring feature, while the ones using scripts could not. This is in line with previous research showing that the students have a good perception of the rubric usefulness (Andrade & Du, 2005; Reynolds-Keefer, 2010). Second, effort could also explain our results. The scripts are cognitively highly demanding (Bannert, 2009; Kostons, van Gog, & Paas, 2009), as their use implies deep reflection and monitoring processes. As for rubrics, they are easier to employ because, when correctly designed, they are concrete and their quality samples are easily compared with the students' final product. Deeper approaches to learning rely on the students' motivation to a greater extent (Kyndt, Dochy, Struyven, & Cascallar, 2010), and even though scripts seem to enhance more self-regulation, they are cognitively more demanding. In sum, students may prefer rubrics as they are easier to use and it is possible to evaluate how the advances relate to the final score.

Limitations

First, one important limitation from this study is that it relies on self-reported data. As previous research has recommended (Boekaerts & Corno, 2005) measurement of self-regulation it is more valid if it has been contrasted with other types of data, mainly process

data (e.g. thinking aloud protocols). Even if we have tried to overcome this flaw by the combination of a general and specific measurement tools (Samuelstuen & Bråten, 2007) this limitation needs to be bear in mind. Second, the sample was mainly comprised by females as usual for pre-service teacher programs. Thus, translation of these results to male students should be done with care, as research has found that in higher education gender plays a role in students' use and strategies of self-regulation (Virtanen & Nevgi, 2010). Third, the use of performance goal as a covariate may have reduced the statistical power of our results. Fourth, the sample size is small (e.g. rubric and script conditions had 20 students) and therefore, like with any other study with small sample sizes, the interpretation of results, in particular confidence intervals and p-values should be done with care.

Educational and theoretical implications

The use of rubrics and scripts seem to have advantages over not using any self-assessment tool for the students' self-regulation. For that reason, their educational use in higher education is strongly recommended. Moreover, in line with previous research (Panadero et al., 2012; Reitmeier & Vrchota, 2009), the use of rubrics is recommended for tasks of low or medium complexity and scripts are recommended for high cognitive demanding tasks. However, the results and limitations of this study have important theoretical and methodological implications. The fact that intervention effects have been found to be more or less effective depending on the assessment tool used points to a methodological weakness that makes it difficult to identify with precision the role of self-assessment tools in self-regulation, self-efficacy and learning. Therefore, intervention effects should be measured systematically controlling the type of task and the instructional conditions. There is also need to explore the effect of "cognitive load", especially in interventions using rubrics as it has never been explored, and such effect might have a major impact on the activation of self-

monitoring (e.g. van Gog, Kester, & Paas, 2011). In conclusion, the use of scripts and rubrics has the potential to enhance students' self-regulation and, therefore, their use is recommended when the conditions for an adequate self-assessment implementation are met (Andrade & Valtcheva, 2009).

References

- Alonso-Tapia, J., Huertas, J. A., & Ruiz, M. A. (2010). On the nature of motivational orientations: Implications of assessed goals and gender differences for motivational goal theory. *The Spanish Journal of Psychology*, 13(1), 232-243.
- Alonso-Tapia, J., & Panadero, E. (2010). Effect of self-assessment scripts on self-regulation and learning. *Infancia y Aprendizaje*, 33(3), 385-397.
- Alonso-Tapia, J., Panadero, E. & Ruiz, M. (2012). Development and validity of the Emotion and Motivation Self-regulation Questionnaire (EMSR-Q). *Paper submitted for publication*.
- 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*. New York: Routledge.
- 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(2).
- Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice*, 48(1), 12-19.
- Andrade, H., Wang, X. L., Du, Y., & Akawi, R. L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *Journal of Educational Research*, 102(4), 287-301.

- 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>
- Bannert, M. (2009). Promoting self-regulated learning through prompts. *Zeitschrift Fur Padagogische Psychologie*, 23(2), 139-145. doi: 10.1024/1010-0652.23.2.139
- 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. (1995). How does self-assessment relate to ideas about learning? In D. Boud (Ed.), *Enhancing learning through self-assessment* (pp. 24-35). New York: RoutledgeFalmer.
- Brown, A. (1987). Metacognition, executive control, self-regulation and other more mysterious mechanisms. In F.E. Weinert & R.H. Kluwe (Eds.), *Metacognition, motivation and understanding*. Hillsdale, NJ.: Erlbaum.
- 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(2), 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(3), 331-350.
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, 59(4), 395-430.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2, 130-144.
- Kitsantas, A., & Zimmerman, B. J. (2006). Enhancing self-regulation of practice: The influence of graphing and self-evaluative standards. *Metacognition and Learning*, 201-212. doi: 10.1007/s11409-006-9000-7

- Klassen, R. M., & Chiu, M. M. (2011). The occupational commitment and intention to quit of practicing and pre-service teachers: Influence of self-efficacy, job stress, and teaching context. *Contemporary Educational Psychology*, 36(2), 114-129. doi: 10.1016/j.cedpsych.2011.01.002
- Kostons, D., van Gog, T., & Paas, F. (2009). How do I do? Investigating effects of expertise and performance-process records on self-assessment. *Applied Cognitive Psychology*, 23(9), 1256-1265. doi: 10.1002/acp.1528
- Kramarski, B., & Michalsky, T. (2010). Preparing preservice teachers for self-regulated learning in the context of technological pedagogical content knowledge. *Learning and Instruction*, 20(5), 434-447. doi: 10.1016/j.learninstruc.2009.05.003
- Kyndt, E., Dochy, F., Struyven, K., & Cascallar, E. (2010). The perception of workload and task complexity and its influence on students' approaches to learning: A study in higher education. *European Journal of Psychology of Education*, 1-23. doi: 10.1007/s10212-010-0053-2
- Heikkilä, A., & Lonka, K. (2006). Studying in higher education: Students' approaches to learning, self-regulation, and cognitive strategies. *Studies in Higher Education*, 31, 99-117. doi: 10.1080/03075070500392433
- Nückles, M., Hübner, S., & Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. *Learning and Instruction*, 19(3), 259-271. doi: 10.1016/j.learninstruc.2008.05.002
- 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. (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., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9(0), 129-144. doi: <http://dx.doi.org/10.1016/j.edurev.2013.01.002>
- 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. doi: 10.1016/j.lindif.2012.04.007
- Peters, E. E., & Kitsantas, A. (2010). Self-regulation of student epistemic thinking in science: The role of metacognitive prompts. *Educational Psychology*, 30(1), 27-52.
- Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, 16(4), 385-407.
- Puustinen, M., & Pulkkinen, L. (2001). Models of self-regulated learning: A review. *Scandinavian Journal of Educational Research*, 45(3), 269-286. doi: 10.1080/00313830120074206
- Reitmeier, C. A., & Vrchota, D. A. (2009). Self-assessment of oral communication presentations in food science and nutrition. *Journal of Food Science Education*, 8(4), 88-92.
- 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>

- Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment Research & Evaluation*, 11. Retrieved from <http://pareonline.net/getvn.asp?v=11&n=10>
- Ruiz-Maya, L. (1977). *Métodos estadísticos de investigación: Introducción al análisis de la varianza*. Madrid: Presidencia del Gobierno, Instituto Nacional de Estadística.
- 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. doi: 10.1348/000709906x106147
- Taras, M. (2010). Student self-assessment: Processes and consequences. *Teaching in Higher Education*, 15(2), 199-209. doi: 10.1080/13562511003620027
- Torenbeek, M., Jansen, E., & Hofman, A. (2010). The effect of the fit between secondary and university education on first-year student achievement. *Studies in Higher Education*, 35(6), 659-675.
- van Dinther, M., Dochy, F., & Segers, M. (2010). Factors affecting students' self-efficacy in higher education. *Educational Research Review*, 6(2), 95-108. doi: 10.1016/j.edurev.2010.10.003
- van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology*, 36(3), 212-218. doi: 10.1016/j.cedpsych.2010.10.004
- Virtanen, P., & Nevgi, A. (2010). Disciplinary and gender differences among higher education students in self-regulated learning strategies. *Educational Psychology*, 30(3), 323-347. doi: 10.1080/01443411003606391

- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated engagement in learning. In D. Hacker, J. Dunlosky & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277-304). Hillsdale, NJ: Erlbaum.
- 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. 13-40).
- 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: Routledge.

Total number of words: 7833.