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

Ernesto Panadero [1], Jesús Alonso-Tapia [2], and Eloísa Reche [3]

[1] *Learning and Educational Technology Research Unit. Faculty of Education. University of Oulu, Finland.*

[2] *Departamento de Psicología Clínica y de la Salud, Universidad Autónoma de Madrid, Spain*

[3] *Departamento Pedagogía, Escuela Universitaria de Magisterio Sagrado Corazón, Córdoba, Spain*

Recommended citation:

Panadero, E., Alonso-Tapia, J., & Reche, E. (2013). Rubrics vs. self-assessment scripts effect on self-regulation, performance and self-efficacy in pre-service teachers. *Studies In Educational Evaluation*, *39*(3), 125-132. doi:10.1016/j.stueduc.2013.04.001

Correspondence concerning this article should be addressed to: Ernesto Panadero, Department of Educational Sciences and Teacher Education, Learning and Educational Technology Research Unit (LET). PO BOX 2000, Fin-90014 University of Oulu (Finland).

E-mail: ernesto.panadero.uam@gmail.com

Acknowledgements: Support for this research was provided by grants from the Spanish Education Ministry to Ernesto Panadero (Ref. SEJ2005-00994) and to Jesús Alonso-Tapia (EDU2009-11765).

Abstract

Two approaches to self-assessment are optimal, because they include 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. A total of 69 pre-service teachers participated in the study. During a semester the participants were trained to design multimedia material in three experimental conditions (rubrics, scripts and control). Results showed that students using the scripts had higher levels of learning self-regulation after the intervention, whereas rubrics decreased performance/avoidance self-regulation (negative self-regulatory actions detrimental to learning). No significant effects were found for students' performance or self-efficacy. Students preferred the use of rubrics to the use of scripts.

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

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

In recent years self-assessment of learning has received a lot of attention, becoming a growing field in educational psychology (e.g. Dochy, Segers, & Sluijsmans, 1999; Ross, 2006; Taras, 2010). The reason for this is that self-assessment is a necessary process 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 students' self-assessment. There are two self-assessment tools being studied to test their potential effects and conditions for effectiveness: rubrics and scripts (Alonso-Tapia & Panadero, 2010). Rubrics are designed to evaluate, mainly but not exclusively, 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 (Bannert, 2009; Kramarski & Michalsky, 2010; Nückles, Hübner, & Renkl, 2009). These two tools have proved to have 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 conclusive, and no prior research has compared their relative effects on self-regulation and performance in a real classroom setting (Panadero, 2011). Hence, this will be the main objective of this study, which was tested in higher education with pre-service teachers.

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 past two decades, as they are crucial competences for higher education students to develop in the transition from secondary education (Torenbeek, Jansen, & Hofman, 2010) and

in order to be successful during university training (Heikkiläa & 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, students must self-assess their ongoing cognitive, emotional, motivational and behavioural processes. By 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 by different names. There are also empirical findings that demonstrate the validity of this connection (Andrade & Valtcheva, 2009; Bannert, 2009; Heikkiläa & Lonka, 2006; Panadero, Alonso-Tapia, & Huertas, 2012): when students self-assess their learning using adequate criteria, they self-regulate with success.

Conditions for adequate self-assessment

However, what implies being able to self-assess one's own learning activity adequately? There is actually a list of conditions for an adequate implementation of self-assessment so that students can learn this skill (Andrade & Valtcheva, 2009). From these, it can be extracted that for appropriate self-assessment to occur, two factors are crucial: (a) using adequate assessment criteria, and (b) using them at the right time (Goodrich, 1996; Panadero, 2011). Therefore, the questions to answer are: (a) what promotes the use of adequate assessment criteria? (b) when is it opportune to use that criteria?

Assessment criteria are the standards against which the execution and the final outcome of a task are evaluated. Although 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). Rubrics and scripts contain these assessment criteria.

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 in the final phase – the self-reflection phase (Zimmerman & Moylan, 2009) – where the students analyse what they have done and reflect on its consequences. Nevertheless, in line with other researchers (Boud, 1995; Winne & Hadwin, 1998), we consider that students cannot only self-assess the final product; they must also consider the process through which the final product is reached. In fact, it can be concluded from previous self-assessment research (Andrade, 2010; Boud, 1995) that a good implementation of self-assessment should influence all the self-regulatory phases – i.e. forethought, execution and self-reflection (Zimmerman & Moylan, 2009). This theoretical perspective is supported by research on the effects of self-regulation interventions: the biggest effects occur when the interventions focus on planning and monitoring or planning and evaluation (Dignath, Büttner, & Langfeldt, 2008). Therefore, instructional help to learn how to self-assess should be given 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 students to evaluate and score their work

without the use of a specific tool. Research has shown that this is not an optimal pedagogical approach, as it is flawed (Dochy, Segers, & Sluijsmans, 1999; Falchikov & Boud, 1989). In this category can be included those interventions that are aimed at enhancing self-assessment, but do not provide students with assessment criteria. As these two approaches do not include assessment criteria, they do not promote precise self-assessment (Andrade & Valtcheva, 2009).

Rubrics are self-assessment tools with three characteristics: a list of criteria for assessing the important goals of a task, a scale for grading the different levels of achievement and a description for each qualitative level (Andrade & Valtcheva, 2009). Rubrics have been 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). Nevertheless, there is a need for more empirical evidence on their direct effect on self-regulation.

Scripts, including cues and prompts, are specific sets of steps structured accordingly to the expert model of performing a task from beginning to end. Like rubrics, they also have positive effects, promoting self-regulation and learning (e.g. Bannert, 2009; Peters & Kitsantas, 2010). Scripts have been used mainly in experimental settings, with only a small number of studies carried out in real settings (e.g. Kramarski & Michalsky, 2010).

As rubrics and scripts contain assessment criteria, they are more effective methods than self-evaluation or self-assessment without assessment criteria. However, what are the main differences between these two tools?

Rubrics and scripts differences

There are two main differences. First, rubrics have a scoring feature that scripts do not. Thus rubrics can emphasize grades, whereas scripts do not have such a characteristic. Second, rubrics include a set of text samples describing characteristics for every performance level, and thus might centre students' attention on outcomes and learning products. On the other hand, scripts are formulated as questions pointing to the steps that the students have to follow, and thus might centre the students' attention on the learning process – in fact, research on scripts shows that they are cognitively demanding (e.g. Kollar, Fischer, & Slotta, 2007). In sum, both tools are oriented towards promoting students' self-assessment, but they present salient different features that can influence different effects.

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 scored higher than the control group on self-regulation and learning, and that the use of the script enhanced self-regulation more than use of the rubric. However, this study was conducted in an experimental setting with secondary education students. Also learning was assessed through a task that was not scored; hence, it remained to test whether intervention effects would be similar in natural classroom settings, with higher education students, and when the task to be performed counted towards the course grades. That is the aim of the present study.

Rubrics vs. scripts effects: research questions

When planning this study, a crucial question was: can different effects be expected from the use of rubrics and scripts in natural classroom settings? Below, we consider this study's expected effects on the dependent variables (self-regulation, performance, self-efficacy and tool's perceived help).

First, it can be expected that *self-assessment will positively affect self-regulation.* However, this effect will depend on the measurement method used. Boekaerts and Corno (2005) recommended using contextual measures of self-regulation, rather than a general self-regulation questionnaire, to evaluate specific intervention effects. However, it can be difficult and costly to measure individual self-regulation in natural classroom contexts. 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). For this reason, in the present study, both a general and a specific self-regulation questionnaire will be used. The instruments chosen for this study measure two types of regulatory actions: (a) learning self-regulation (regulatory actions oriented to learning goals) and (b) performance/avoidance self-regulation (regulatory actions oriented to performance and avoidance goals). This differentiation is based on Boekaert's top-down and bottom-up model (Boekaerts & Corno, 2005). With regard to *learning self-regulation*, previous research shows that both tools enhance the use of these strategies (Bannert, 2009; Panadero & Jonsson, 2013). Regarding *performance/avoidance self-regulation*, the self-assessment tools should decrease this type of negative self-regulation, as they give the students clear criteria for performing the task.

Second, though *performance* does not depend exclusively on self-assessment and self-regulation, these factors are relevant. Therefore, *some improvement in performance can be expected as a result of promoting self-assessment*.

Third, if self-assessment affects self-regulation and performance in a positive way, *it may also produce an improvement in self-efficacy*. Why is self-efficacy important for these study goals? There are two reasons. First, pre-service teachers' self-efficacy has been linked

to their later commitment during their adaptation to the workplace, and to performance (Klassen & Chiu, 2011). Second, and more directly related to this study's goals, self-efficacy has an essential role in self-regulation: if students have low self-efficacy expectations for a task, 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 was the case, the self-efficacy level prior to intervention could mediate the effect of self-assessment in this variable; consequently, self-efficacy needs to be measured before the intervention.

Previous research on the effects of rubrics and scripts on self-efficacy is not conclusive. Alonso-Tapia and Panadero (2010) studied whether scripts enhanced self-efficacy in comparison to a control group based in previous research by Andrade et al. (2009) that studied the same topic using rubrics. Andrade found that rubrics tended to enhance self-efficacy beliefs in interaction with gender (more so for girls than for boys). Alonso-Tapia and Panadero (2010) found no significant effect in self-efficacy for the use of scripts, attributing this result to the short length of the intervention. In conclusion, the question of the potential effects of scripts on self-efficacy remains open, as does the question of which tool might promote a higher increase in self-efficacy.

Fourth, to be motivated to use rubrics and scripts, *students need to perceive these tools as useful*. Previous research shows that students value the use of rubrics (Andrade & Du, 2005; Reynolds-Keefer, 2010), but there is no empirical evidence about this topic in relation to scripts. Therefore it remains unclear which tool would be perceived as more useful by the students.

In sum, the *research questions* and *hypotheses* of this study are thus as follows:

(a) Do rubrics and scripts enhance self-regulation when compared 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 more in comparison to the rubric group because scripts are more cognitively demanding (H1b). Regarding performance/avoidance self-regulation, both self-assessment tools' groups would show lower levels of this type of negative self-regulation than the control group, but there is no hypothesis on the comparison between rubric and scripts (H1c).

(b) Do rubrics and scripts enhance performance over the control group? Both tools will have positive effects on performance compared to the control group as far as they have positive effects on self-regulation (H2).

(c) Do rubrics and scripts enhance self-efficacy in comparison with the control group? It is expected that the self-assessment tools will increase self-efficacy in comparison to the control group (H3a), but it is unclear which tool will produce a bigger effect (H3b).

(d) Which self-assessment tool – rubric or script – do students perceive as the better learning tool? There is no hypothesis on the direction of this effect (H4).

Method

Participants

The sample was comprised of 69 participants from three classroom groups: 20 in the rubric condition (29%), 20 in the script condition (29%) and 29 in the control group (42%). Each classroom group was assigned randomly to each experimental condition. The majority of the participants were females -usual in pre-service teacher programs-: 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 enrolled in the course “New technologies applied to education”. This course belongs to the second year of the teacher training program. The main goal of the course is to prepare future teachers for the use of new technologies with pedagogical purposes. Consequently, students have to learn how to design multimedia-learning material. This context was selected because the students were going to acquire new skills and the intervention effects would have been less affected by prior knowledge.

The students had been randomly assigned to their natural classroom groups by the university administration, and that same grouping was used to form the experimental conditions. Even if students were randomly assigned to their groups, the fact that this was not controlled by the researchers 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 is a predictor of self-regulation because depending on the students’ goals they will activate the strategies needed to self-regulate (Pintrich, 2004). For this reason, goal orientation was measured to confirm whether the three experimental conditions showed similar scores on goal orientations. In differences were found goal orientation would be used as a covariate in the analyses.

Materials

Instruments for assessing dependent

Self-regulation measures. In order to reach a good estimation of self-regulation, following the ideas by Samuelstuen and Bråten (2007), two different measures were used for assessing self-regulatory processes: a general and a specific questionnaire.

General self-regulation questionnaire: Emotion and Motivation Self-regulation Questionnaire (EMSR-Q) (Alonso-Tapia, Panadero & Ruiz, 2013). 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 a 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 goals or to avoid the task, for example: “I am getting nervous. I don’t know how to do it”. The higher the value in this scale, the more negative actions detrimental for learning occur.

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 of them were discharged after the factor analysis), α = .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 of them were discharged after principal component factor analysis, as they loaded below .30, the standard level chosen for including or not an item in the scale), α = .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 sample 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.

Performance. To measure performance, the Multimedia Material and Web Quest/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 Kappa = .89.

Self-efficacy questionnaire. It was created for this study to measure the students' self-efficacy towards the activities they were going to perform in the course. It is compounded by three different scales based on the different task the students had to perform: (1) *Didactic Unit Design scale* (six items, Cronbach α = .86), that assesses the student's self-efficacy for designing a didactic unit; (2) *Multimedia Material Design scale* (six items, α = .86), that assesses self-efficacy for designing PowerPoint presentations; (3) *Web Quest/Treasure Hunt scale* (six items, α = .89), that assesses self-efficacy for designing either of these specific multimedia tasks. Self-efficacy was measured before and after the intervention.

Tool perceived help. Students in the rubric and script groups were asked using an item to be answered in a five points Likert scale how helpful rubrics and scripts were. This was measured after the intervention.

Instruments for control variables

Previous experience. Two five-point Likert items were created to measure previous experience in the design of multimedia material.

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 (α = .88), performance (α = .88), and avoidance goals (α = .83). The items used five-point Likert scales ranging from 'completely disagree' to 'completely agree'.

Instruments used for the intervention

Self-assessment tools: Rubrics and scripts. Expert samples of PowerPoint presentations and Web Quests/Treasure Hunts were analyzed to create the rubrics and scripts. After extracting the assessment criteria from these samples, two rubrics and two scripts, one for each task, were created by the three researchers. Rubrics were used by the teacher and one of the researchers to evaluate the students' performance (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 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 and modeled how the tools 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 developed 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, which were mainly based in a short introduction by the teacher and then individual work. The teacher and one of the researchers had a meeting every two weeks to monitor the progress of the course and discuss any possible deviation from the established standard procedure.

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.

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 (e.g. their own score in the rubrics categories). In the last session, they received their work scored and 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 also reported 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 and the WebQuest/Treasure Hunt score. As mentioned earlier, 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. 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 analysis with the dependent variables.

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 (specific self-regulation, performance, and tool perceived help). Main effects and interactions were tested, and when necessary post-hoc Bonferroni tests were carried out.

Results

Preliminary analyses

Correlations between 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.

Insert Table 1.

Normality tests. All the normality tests were non-significant, though "Perceived Ability for WebQuest/Treasure Hunt" had a $p = .053$, a value close to 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, SD = 4.87) were significantly less oriented to performance than rubric (M = 53.87, SD = 4.79) and script (M = 52.24, SD = 4.20) students, F (2, 58) = 4.78; $p < .05$. Thus performance orientation was used as co-variable for the remaining analyses.

Previous experience. The three conditions did not differ on their levels of previous experience on the course tasks they were going to perform (p = .65). Therefore this variable was not used as co-variable.

Intervention effects on self-regulation

Self-regulation was measured using two questionnaires: 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 EMSR-Q was used twice before and after the intervention and the effect of 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 groups 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 had 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; η^2 = .051; Rubric Adjusted M = 20.87, SD = 4.24, Script Ad. M = 20.93, SD = 3.55, Control Ad. M = 18.43, SD = 4.24, or on the WebQuest/Treasure Hunt scale, F (2, 69) = .866; p = .426; η^2 = .031; Rubric Adjusted M = 24.93, SD = 4.83, Script Ad. M = 24.27, SD = 4.64, Control Ad. M = 22.04, SD = 5.29.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 self-regulation –which is detrimental for learning- came from the EMSR-Q performance/avoidance scale. The interaction between occasion and training was significant, F (2, 66) = 5.72; $p < .001$; η^2 = .175. Differences are illustrated in Figure 2, post-hoc Bonferroni analysis showed that this effect is caused by the rubric-group score decrease ($p < .001$) whereas the other groups remain at 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 was null for 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 hypothesis (H2).

Intervention effects on self-efficacy

Self-efficacy was measured before and after the intervention through three specific scales. It was found differences between groups only for the Multimedia Material scale F (1, 53) = 3.871; $p < .05$, $\eta^2 = .127$; Rubric Adjusted M = 19.16, SD = 4.26, Script Ad. M = 20.33, SD = 3.13, Control Ad. M = 17.4, SD = 4.79. 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. Therefore the hypothesis about the bigger gain in comparison with the control group has to be rejected (H3a) and it cannot be concluded what tool produces the bigger effect (H3b).

Tool perceived help

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, F (1, 28) = 22.76; $p < .001$; $\eta^2 = .477$; Rubric Adjusted M = 4.7, SD = .439, Script Ad. M = 3.8, SD = .561. The direction of this result was not hypothesized (H4).

Discussion

The aim of this study was to test the effects of different self-assessment tools (rubric and script) in university students' self-regulation, performance and self-efficacy in a natural context. It also explored which tool the students preferred.

Self-regulation

While scripts enhanced learning self-regulation, rubrics decreased performance/avoidance self-regulation, highlighting that these instruments have different effects on the students. Thus, the study results are interesting for future research, because they provide information about aspects that need to be clarified and that can be used to determine different uses of rubrics and scripts.

Regarding learning self-regulation, this study *partially* supports the notion that scripts increase 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 metacognitive awareness and therefore to activate more learning strategies. We write *partially* because this idea is supported only by the results from the general self-regulation questionnaire, not those from the specific questionnaire. However, the latest questionnaire results can be affected by the items content. That questionnaire refered to specific regulatory actions about the tasks and students from the three conditions might have performed those actions regardless of the use of rubrics or scripts. For example, one of the items in that questionnaire is “Are the steps to reach the end of the WebQuest/Treasure Hunt clearly defined?”, and the steps needed to be clearly defined to get the approval from the teacher to submit the WebQuest/Treasure Hunt final version. Thus, it could be difficult to draw any general conclusion about the effect on learning self-regulation based on this study alone.

Nevertheless, our results should be interpreted in the context of similar studies’ results. In a previous study, Panadero et al. (2012) used thinking-aloud protocols to measure learning self-regulation, comparing the effect of rubrics, scripts and control in an experimental setting

with secondary students, and found that scripts had the highest positive effect on learning self-regulation, followed by rubrics, in comparison to the control group. Consequently, the partial results found here, and the stronger ones coming from similar research using a more objective measurement of self-regulation, highlight 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 the effect of self-assessment on this type of self-regulation. Panadero et al. (2012) used 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 grade, in the present study the students' performance was critical for their final grade. Therefore, the participants in this study could have felt more pressure to perform optimally because they had a grade pending on their work. For that reason, rubrics might have made a difference, as they clarify how the students' performance will correspond with the instructor's grade. On the other hand, scripts are more cognitively demanding – that is, the students need to activate more strategies to use the scripts properly – but do not offer information about how the performance relates to the final score. For these reasons, rubrics might be better tools in tasks that will be graded, but this hypothesis should be tested in future research.

In conclusion, scripts seem to have the potential to activate students' use of learning strategies, while rubrics seem to have the potential to decrease the use of negative self-regulation strategies. One possible implication of these results may be that for activities in

which the students might be experiencing anxiety (e.g. activities that have a deadline and will be scored), rubrics could help students to focus more on the learning process; scripts, meanwhile, might have an advantage if used in complex activities that require deep processing, guiding the students to activate a greater number of learning strategies (Reitmeier & Vrchota, 2009).

Performance

Two related matters can explain the self-assessment instruments' lack of effects on performance. First, performance was measured through the course final grade, which might have been affected by factors beyond the use of self-assessment tools that were not controlled, even if we had offered equivalent pedagogical environments in the three experimental conditions (same teacher, same instructions, same agenda, etc.). Second, the structure of the activities was highly defined, and the control group may have benefited from having these highly defined tasks to perform. Therefore, the effects of the self-assessment tools could 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 (e.g. Kramarski & Michalsky, 2010; Peters & Kitsantas, 2010) have positive effects on performance and learning. Therefore the lack of effects in this study can be explained by the type and nature of the task chosen, as will be explained in more detail later.

Self-efficacy

There were no significant differences between the self-assessment tools' groups and the control in relation to self-efficacy. An explanation would be that when students perform a

task and obtain a final product, they might experience that “they are able” to do that task. Therefore, it seems plausible that all students in our study maintained their perception of efficacy, as all of them created the final products: a didactic PowerPoint presentation and a Web Quest/treasure hunt.

Under what conditions, then, could self-efficacy be enhanced? Two complimentary conditions may be considered. First, it may be that increased self-efficacy depends on feedback, and in particular its frequency and characteristics. In the study by Panadero et al. (2012), students in the different conditions received feedback on three occasions; all the experimental groups increased their self-efficacy, but more so if feedback was provided on process rather than on performance. Besides, a significant interaction was found between type of feedback and self-assessment tool: the group that worked with rubrics and received mastery feedback saw the greatest increase in self-efficacy. Therefore, the existence and type of feedback may explain changes in self-efficacy. In the present study, the scarcity of feedback – it was given only once – and its nature – mainly 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 experience – 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, will self-efficacy improve. 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 regarding self-efficacy. For these reasons, future work should study the effect of rubrics and scripts on self-efficacy using a stronger measure than that used here.

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

Rubrics and scripts had a positive effect on self-regulation, but not on performance and self-efficacy. This fact puts into question the relations between these variables which is actually strongly documented in previous research: self-efficacy and self-regulation have an impact on performance (e.g. Zimmerman & Schunk, 2011). However, from our results 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 lead to better performance if the task can be performed more efficiently with less reflection, as not all tasks need the same level of metacognitive activity to produce positive outcomes (e.g. 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 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. Therefore, the lack of effects on performance and self-efficacy may have been due to both task choice and the pedagogical setting, plus, in the case of self-efficacy, the scarcity of feedback. Consequently, it is necessary that future research conducts a systematic study of these variables' effects, taking into account the task and instructional context properties.

Tool's perceived level of help

The fact that the rubrics were perceived as more helpful than the scripts may have two complementary explanations. First, rubrics could be promoting a feeling of security based on the scoring feature, as explained before. It is known that 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 the instructor's grade using the rubric scoring feature, whereas the students using scripts could not. This is in line with previous research showing that students have a good perception of the use of rubrics (Andrade & Du, 2005; Reynolds-Keefer, 2010). Second, effort could also explain our results. The scripts are cognitively highly demanding, as their use implies metacognitive awareness and monitoring processes (Bannert, 2009; Kollar, Fischer, & Slotta, 2007; Kostons, van Gog, & Paas, 2009). Rubrics are easier to employ because, when correctly designed, they are concrete and their quality samples are easily compared to the 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 learning self-regulation more, they are cognitively more demanding and, therefore, less motivating for students. In sum, students may prefer rubrics as they are easier to use and can help the students evaluate how their progress relates to the instructor's score.

Limitations

One important limitation of this study is its reliance on self-reported data. As previous research has noted (Boekaerts & Corno, 2005), measurement of self-regulation is more valid if it has been contrasted with different types of data, mainly process data (e.g. thinking-aloud protocols). Although we have tried to overcome this flaw by combining a general and a specific questionnaire (Samuelstuen & Bråten, 2007), this limitation needs to be borne in

mind. Second, the sample mainly comprised females, as is common for pre-service teacher programmes. 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 of self-regulatory strategies (Virtanen & Nevgi, 2010). Third, the sample size is small (the rubric and script conditions had 20 students) and therefore, as with any other study with small sample sizes, interpretation of the results should be cautious – particularly in relation to confidence intervals and p-values. Fourth, the use of performance goal as a covariate may have reduced the statistical power of our results.

Educational and theoretical implications

The use of rubrics and scripts presents advantages for students' self-regulation in comparison with the use of no self-assessment tool. For this 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, whereas scripts are recommended for highly cognitively demanding tasks. In addition, this study's results and limitations 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 performance. Therefore, intervention effects should be measured systematically, controlling the type of task and the instructional conditions. Nevertheless, the use of scripts and rubrics has the potential to promote students' self-regulation, and therefore their use is recommended when the conditions for an adequate implementation of self-assessment 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. Kluve (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.

Goodrich, H. W. (1996). *Student self-assessment: At the intersection of metacognition and authentic assessment*. (57), ProQuest Information & Learning, US.

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

Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*(6), 708-721. doi: 10.1016/j.learninstruc.2007.09.021

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

Heikkiläa, 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.

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

Table 1. Correlations between pre-intervention measured variables

	PO	AO	PEM	PEWQ	PSR	LSR	PAMD	PAWD
LO	-,413**	-,538**	,327*	-,074	-,256	,112	,191	,087
PO		,285*	-,061	,119	,265*	,251	,075	-,107
AO			-,088	-,026	,267*	-,013	-,162	-,275*
PEXP_MMM				-,010	-,272*	,048	,024	,042
PEXP_WQ					,150	-,068	-,046	,043
PSR						-,362**	-,093	-,273*
LSR							,164	,090
PA_DUD								, 417**

** p < .01, * p < .05

LO: Learning goal orientation; PO: Performance goal orientation; AO: Avoidance goal orientation; PEM: Previous experience with multimedia design: PEWQ: Previous experience with Web-Quest design; PSR: Performance oriented self-regulation; LSR: Learning oriented self-regulation; PAMD: Perceived ability in multimedia design; PAWD: Perceived ability in web-Quest/Treasure-Hunt design.

Figure 1. Interaction effect of the self-assessment instrument per the occasion on Learning SR based on the adjusted means.

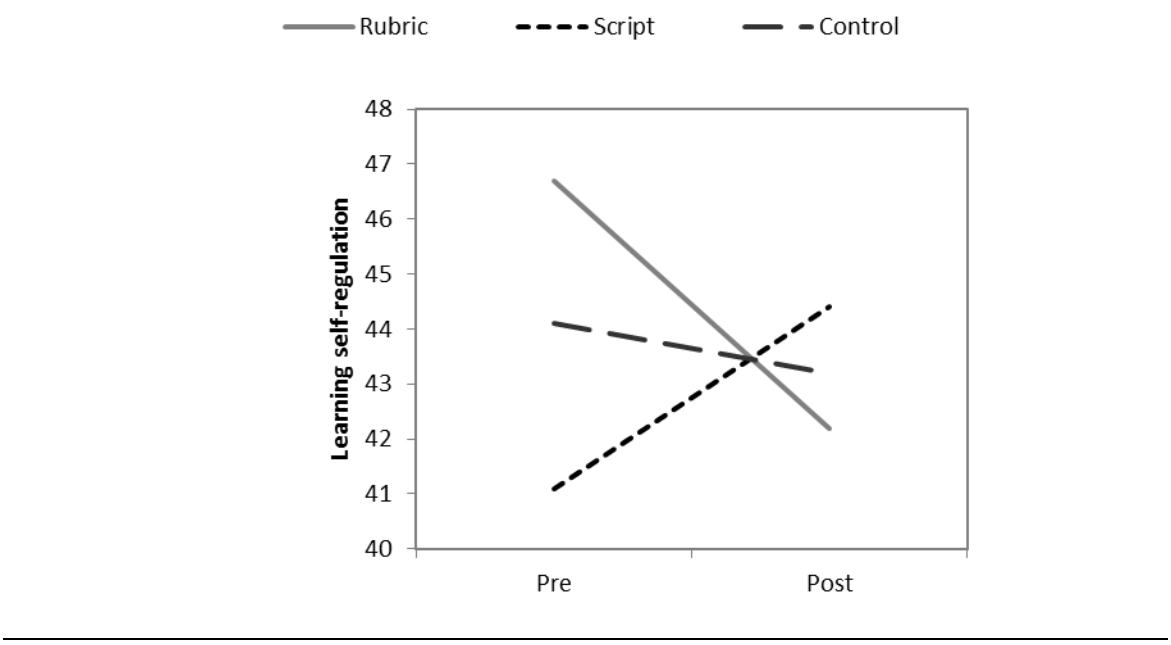

Figure 2. Interaction effect of the self-assessment instrument per the occasion on Performance/Av. SR based on the adjusted means.