

The impact of rubrics and scripts on self-regulation, self-efficacy and performance in collaborative problem-solving tasks

Juan Fraile¹, María Gil-Izquierdo², and Eva Medina-Moral²

¹Universidad Francisco de Vitoria, Madrid, Spain

²Department of Applied Economics, Universidad Autónoma de Madrid, Madrid, Spain

Recommended citation:

Fraile, J., Gil-Izquierdo, M., & Medina-Moral, E. (2023). The impact of rubrics and scripts on self-regulation, self-efficacy and performance in collaborative problem-solving tasks. *Assessment & Evaluation in Higher Education*, 48(8), 1223-1239. <https://doi.org/10.1080/02602938.2023.2236335>

Author note

Juan Fraile <https://orcid.org/0000-0002-7212-7380>

María Gil-Izquierdo <https://orcid.org/0000-0001-8268-575X>

Eva Medina-Moral <https://orcid.org/0000-0001-9073-0223>

Funding: (1) Universidad Francisco de Vitoria, Call for Research Educational Innovation 2020, Reference number UFV2020-46. (2) Universidad Autónoma de Madrid, Call Teaching Innovation Projects 2017/18, Reference number CEE_018.17_IMP. (3) Spanish National R+D call from the Ministerio de Ciencia, Innovación y Universidades (Generación del conocimiento 2020), reference number: PID2019-108982GB-I00.

Correspondence concerning this article should be addressed to Juan Fraile, Universidad Francisco de Vitoria, Carretera Pozuelo-Majadahonda km. 1.8 – 28223 Pozuelo de Alarcón, Madrid, Spain. Email: juan.fraile@ufv.es

Abstract

Rubrics and scripts are tools that can promote self-assessment. Although similar, rubrics and scripts differ in the metacognitive processes they support, their cognitive demands, how students use them, and their focus. While there is ample evidence of the positive effects of rubrics, there is less research on the impact of scripts and their use in collaborative tasks. We aimed to compare the effects of rubrics and scripts on self-regulation, self-efficacy, academic performance, and students' perceptions when used in group tasks on econometrics. A total of 134 university students participated in a quasi-experiment under three conditions (rubrics, scripts, control). Positive effects on self-regulation were found for the experimental groups, and a positive impact on self-efficacy only for the rubric group. Additionally, as previous academic performance differed among conditions, difference-in-differences analysis was employed. In the rubric condition, the students that started from very high levels of academic achievement benefited more from using the rubric. In contrast, in the script condition, it was the students with intermediate or high levels of academic performance who experienced the greatest benefits from employing the script.

Keywords: rubric, script, self-assessment, self-regulated learning

The impact of rubrics and scripts on self-regulation, self-efficacy and performance in collaborative problem-solving tasks

Self-assessment is a practice that has been intensively researched, although its complexity and multiple ways of conceptualising and implementing it imply that there are still many open lines of research on how and why it works (Panadero et al., 2016; Andrade, 2019). Importantly, self-assessment is a crucial process in the self-regulated learning (SRL) cycle, especially in the self-reflection phase (Andrade & Valtcheva, 2009). This relationship is attested via both empirical research that shows the direct effect of self-assessment on self-regulated learning (Panadero et al., 2017) and academic achievement (Yan et al., 2022). Since self-assessment is such a crucial educational process, different tools have been created to promote its enhancement, such as rubrics and scripts.

With regard to rubrics, there is a large and growing body of evidence on their use (e.g. Dawson, 2017) and their positive formative effects (e.g. Panadero & Jonsson, 2013; Brookhart, 2018). On the other hand, research on the effects of scripts is more limited. Both instruments, although similar, are different in terms of the metacognitive processes they drive, their cognitive demands, their focus, and the way they are used by students.

Unfortunately, to the best of our knowledge, there have been only a few studies (Panadero et al., 2012, 2013, 2014) that have examined the comparative effectiveness of rubrics and scripts. Although these studies mainly found positive effects on self-regulated learning, and mixed results on self-efficacy and performance, it is important to note that further research is still needed, particularly with regards to different tasks. The designs employed in these studies presented certain characteristics that warrant further investigation, such as their implementation in a laboratory context (2012), limited participant enrolment (2013), or the absence of a control

group (2014). Furthermore, in these studies, students performed a landscape analysis (2012), designed multimedia materials and summarised texts (2013), and created conceptual maps (2014). In contrast, our study focused on three tasks involving challenging econometric problems. Therefore, this study seeks to provide additional evidence on the use of rubrics and scripts, particularly in the context of group tasks involving complex econometric problem solving.

Rubrics and scripts

Definitions

A rubric is a tool that consists of three characteristics: a list of criteria; performance quality levels; and a description for each performance level for each criterion. In addition, a grading system is usually associated with the rubric to generate a score (Andrade & Valtcheva, 2009).

Scripts are a list of steps from beginning to end to help the students perform or solve a given task (Wood et al., 1976). Usually presented as consecutive questions that students need to answer, the script's purpose is to guide the learner through the steps to solve the task as an expert would do (Panadero et al., 2016).

Differences among rubrics and scripts

In terms of grading, rubrics provide scoring strategies and a focus on the final product, providing students with clarity about teacher expectations and scoring, whereas scripts prioritise the process (Panadero et al., 2014). With regard to comprehension and cognitive demand, while it has been claimed that scripts are an easy tool for learners to use, require few steps, do not increase task performance time, and improve transfer (Atkinson et al., 2003), Kollar et al. (2007) highlighted their cognitively demanding nature. Brookhart (2018) contends that rubrics, while

analytical, require an extensive and nuanced process, including the use of other strategies such as exemplars and self- and peer-assessment, to effectively understand learning objectives. Taking into account all of the above elements, the available evidence that we are aware of suggests that rubrics are more beneficial in emphasising learning objectives than in focusing on the grade (Panadero et al., 2014). In terms of self-regulatory strategies, a comparative study between rubrics and scripts showed that scripts led to a higher use of such strategies, with similar academic performance results (Panadero et al., 2012). This may indicate a potentially higher cognitive load associated with scripts.

Effects of rubrics vs scripts

With regard to SRL, this refers to “self-generated thoughts, feelings, and actions that are planned and cyclically adapted to the attainment of personal goals” (Zimmerman, 2000 p. 14). According to Zimmerman, SRL is composed of three phases (forethought, performance, and self-reflection). Rubrics, by showing how the outcome of the task should look, encourage SRL in all its phases, although some have argued that they focus on the third phase (Brookhart, 2018) but students find they can help in all phases (Panadero & Romero, 2014). The scripts, however, with their step-by-step design, can encourage the learner to self-assess at each step of the process.

Previous research comparing both tools showed mixed results. In Panadero et al. (2012), both rubrics and scripts improved students’ self-regulation, measured through think-aloud protocols compared to a control group and, moreover, in the case of scripts an even greater effect was obtained. In two other studies (Panadero et al., 2013, 2014), the authors employed the EMSR-Q questionnaire (Alonso-Tapia et al., 2014). In both studies, the use of rubrics decreased avoidance self-regulation, which is related to actions derived from negative emotions and whose

decrease would lead to an increase in learning. In addition, the use of scripts increased learning self-regulation, related to mistake correction.

Regarding self-efficacy, a meta-analysis (Panadero et al., 2017) shows that, contrary to expectations, students who used self-assessment tools (i.e. rubrics, scripts) reported lower self-efficacy than those who did not. The hypothesis argued by those authors is that students become aware of the complexity of high-quality performance as a result of using these instruments. In previous studies in which scripts and rubrics were indeed used and compared, no significant differences were found (Panadero et al., 2013, 2014), or were present in only a few combinations together with the provision of feedback (Panadero et al., 2012).

Concerning academic performance, using rubrics under the guidelines of formative assessment and in combination with metacognitive activities, has generally been shown to increase academic performance (see Panadero & Jonsson (2013) for a review). The use of scripts has also reflected positive results compared to using no tool at all (e.g. Alonso-Tapia & Panadero, 2010). In previous studies in which scripts and rubrics were compared (Panadero et al., 2012, 2013, 2014), the results were heterogeneous.

Research on students' perceptions of rubrics point to the fact that it helps them understand teacher expectations (Reynolds-Keefer, 2010; Jonsson, 2014). Students' impressions of the utility of rubrics and scripts were measured comparatively by Panadero et al. (2014), who found that students who used rubrics reported a higher usefulness of this instrument compared to those who used scripts. Furthermore, students who used rubrics reported to be more focused on learning than students who used scripts, and there were no significant differences between the groups when it came to focusing on grades.

Aim and research questions (RQ)

Our aim is to explore the effects of rubrics and scripts on third-year higher education business students' self-regulation (RQ1), self-efficacy (RQ2), performance (RQ3), and perceptions about these instruments and their use (RQ4).

Method

Research design

This study has a quasi-experimental design with three conditions (rubrics vs scripts vs control). Two dependent variables were measured before and after the intervention – i.e. self-regulated learning and self-efficacy – and the three other dependent variables just after the intervention: academic performance (three activities); perceptions; and reported use of rubrics or scripts.

Participants

The sample consisted of 134 participants from three intact classroom sections with the same teacher (third author). Students were randomly assigned to each section at the beginning of the programme. Forty-four participants were in the rubric condition (32.8%), 45 in the script condition (33.6%), and 45 in the control condition (33.6%). The majority of participants were male (57.46%) with a mean age of 21 years ($SD = 1.79$). The students were enrolled in the “Econometrics II” course of the third year of the Business Administration and Management degree, which is a complex and challenging course. Participation in the study was voluntary, although it was integrated into the instructional design of the course. In the second session, prior to data collection through the questionnaires, the students were informed about the research and given the information sheet and the informed consent form. All the students agreed to participate. The present study was conducted in accordance with the guidelines and regulations set forth by

the Ethics Committee of the first author's affiliated institution, which granted formal approval (14/2023).

Materials

Instruments for assessing dependent variables

Self-regulated learning measures. In order to reach an appropriate estimation of self-regulation, two different instruments were used following prior suggestions (Boekaerts & Corno, 2005; Samuelstuen & Bråten, 2007):

Motivated strategies for learning questionnaire (MSLQ) (Pintrich et al., 1991). We used this questionnaire, the most commonly used to measure SRL (Roth et al., 2016). We selected six scales from the MSLQ, for a total of 37 items with a 5-point Likert scale (“almost never”–“almost always”). Those six scales were chosen because they were the most relevant to our study. The corresponding alpha values pre- and post-intervention were: intrinsic goal orientation (.65–.67); test anxiety (.66–.75); self-efficacy for learning and performance (.79–.87); extrinsic goal orientation (.56–.59); organisation (.80–.70); and metacognitive self-regulation (.72–.68).

Specific self-regulation questionnaire (SSR-Q). This questionnaire was created ad hoc as it is recommended to also include scales to measure specific task regulatory strategies (Samuelstuen & Bråten, 2007). It is composed of seven items with a 5-point Likert scale (“almost never”–“almost always”). The items refer to specific self-regulatory actions related to the activities the students performed. The reliability was $\alpha = .70$.

Self-efficacy questionnaire. Created ad hoc to measure the specific task self-efficacy, partly developed from previous research (e.g. Panadero et al., 2014). It includes five items with a 5-point Likert scale (“almost never”–“almost always”). The items refer to econometric problem

solving and writing reports, which were the activities students carried out. The reliability index was $\alpha = .70$ and $.80$, respectively.

Performance measures. The study comprises five different measurement points for students' performance. The third author was the only teacher and was in charge of the three student sections and graded all the students' tasks and examinations. First, the three group tasks in which the rubrics or scripts were used: (1) Steps in regression analysis; (2) Problems in econometric models: functional forms, multicollinearity and structural stability of regression models; and (3) Problems in econometric model: heteroscedasticity. Second and third, a midterm examination and the final examination of the course.

Students' perceptions, and how students report employing the rubrics/scripts.

Quantitative questionnaire. A questionnaire was created for this study to measure students' perceptions about rubrics/scripts and their use. It included ten 5-point Likert scale ("completely disagree"–"completely agree") items with a reliability index of $\alpha = 0.89$.

Qualitative questionnaire. Six open-ended questions were asked to find the perceived benefits of using the tool (rubrics or scripts, in each case), disadvantages and limitations, ease or difficulty of use, effect on stress or anxiety, and co-creating the tools together with the teacher instead of being provided directly (although in this study the tools were not co-created between teacher and students). The data were coded according to the six categories above – equal to the questions – and into inductive categories within them. Upon completion of the analysis, due to the large number of subcategories (e.g. 41 distinct subcategories on benefits, 46 on disadvantages, and 46 on employment), it was decided to group them into broader subcategories.

Students' use of rubrics/scripts.

Student-reported use. A questionnaire was designed for students to report on their use of the rubrics/scripts. The questionnaire asked about (1) overall use of the tool (1 item), (2) use for each phase of the SRL (Zimmerman, 2002) (3 items), and (3) use for each of the three tasks (3 items). Therefore, it included seven 5-point Likert scale items (“I have not used the rubrics/scripts”–“I have used rubrics/scripts a lot”) with a reliability index of $\alpha = 0.84$.

Instruments used for the intervention

Three rubrics (Supplementary Material A) and three scripts (Supplementary Material B) were employed for the execution of the three assignments. In their design, we tried to make them as similar as possible in terms of the content and sections of the assignment so that the only difference would be the format of the instrument itself.

Data analysis

First, to evaluate the effects of the instruments in the five dependent variables, we used ANOVAs. Repeated measures ANOVAs were used to determine the interaction effect between groups (control vs rubric vs script) and over time (before vs after the intervention) for the three variables, measured before and after the intervention (SRL, SSRL and self-efficacy). Effects of group and time were also examined. Second, due to the differences in previous academic achievement among the conditions, difference-in-differences analysis with multiple time periods (Callaway & Sant’Anna, 2021) was used. This technique allows the measuring of the difference between the pre–post intervention, within the experimental and control groups. That is, the difference-in-differences estimator measures how much the experimental group changes over time with respect to how the control group changes over time. To establish the temporal comparisons in the difference-in-differences analysis, three alternative measures – temporary changes – are used: (1) the final grade of the subject from the previous academic year (period 1)

is compared to the final grade of the course of this research (period 2); (2) the grade of the midterm examination (period 1) is compared to the final grade examination in the course of this research (period 2); (3) the grade of the first group assignment (period 1) is compared to the grade of the third group assignment in the course of this research (period 2).

In addition, students were segmented by quartiles according to the grade obtained in the similar subject in the previous academic year and according to the academic record grade (low-intermediate-high-very high), to explore the effects according to the grade students obtained and to identify which students benefit the most from the use of rubrics or scripts.

Procedure

In preparation for the course prior to its start, two of the authors of this research, specialists in formative assessment, advised the teacher of the course on the use and design of rubrics/scripts.

The second session, in the experimental groups, was used to collect data and explain to the students what the rubrics/scripts are and how to use them. This was done by means of a short exercise. The groups were formed by the students (maximum of five students).

In order to avoid the use of rubrics or scripts in the sections where these tools were not used, these tools were handed out to students in class on paper and were published in the course learning management system (LMS; Moodle), but with the limitation that only the students in each group could access and view the tool that corresponded to them (or none in the case of the control group). Subsequently, in the three group tasks in which the rubrics and scripts were used, the teacher provided these tools on paper and spent a few minutes of the session explaining the task, the tools and leaving time for the groups of students to start planning (first phase of SRL) and carrying out the task.

In the last session of the course, the students completed the questionnaires and, in the case of the experimental groups, reported their perception of the rubrics/scripts and their use of the tool. A few days later the students took the final examination of the course.

Results

To explore whether the conditions were comparable, two actions were performed. First, we compared the academic records of the students from their first and second university years (grade point average), which were close to significance at the 95% confidence level [$F(2, 131) = 2.527, p = .084, \eta^2 = .037; M_{\text{con}} = 6.53, M_{\text{rub}} = 6.48, M_{\text{scr}} = 6.3$]. Second, we compared their grade in “Econometrics I”, which was the first part of their course in which we ran this study, and found significant differences [$F(2, 131) = 13.19, p = .000, \eta^2 = .168; M_{\text{con}} = 7.38, M_{\text{rub}} = 5.77, M_{\text{scr}} = 6.51$]. The control condition outperformed the rubric condition (Dif.= 1.61; $p = .000$) and the script condition (Dif.= .87; $p = .018$). The difference between the scripts group and the rubrics group was .74 and almost significant at the 95% confidence level ($p = .059$). Therefore, the rubric condition had the worst starting level, followed by the script and then the control.

RQ1: What are the effects of rubrics and scripts on self-regulated learning?

Motivated strategies for learning scales (MSLQ)

The repeated measures ANOVA revealed any significant differences between the three groups (control vs rubric vs script) across time (pre-post intervention) [$F(2, 121) = 5.08, p = .008, \eta^2 = .08$]. Although there were no significant differences between the groups with reported similar levels of self-regulation both before [$F(2, 121) = .15, p = .86, \eta^2 = .003; M_{\text{con}} = 91.61, M_{\text{rub}} = 89.98, M_{\text{scr}} = 90.51$] and after the intervention at the 95% confidence level – though we did find at the 90% level [$F(2, 131) = 2.42, p = .09, \eta^2 = .168; M_{\text{con}} = 77.71, M_{\text{rub}} = 86.71, M_{\text{scr}} = 91.0$] – the control group showed a significant decrease after the intervention ($p = .000$).

Specific self-regulation questionnaire (SSR-Q)

There was not a statistically significant difference between the three groups over time [$F(2, 111) = .006, p = .99, \eta^2 = .00$]. However, the occasion was significant [$F(1, 111) = 36.22, p = .000, \eta^2 = .25$] due to the overall increase in specific self-regulation strategies among all conditions before [$M_{\text{con}} = 17.37, SD_{\text{con}} = 2.6; M_{\text{rub}} = 17.13, SD_{\text{rub}} = 3.58; M_{\text{scr}} = 17.26, SD_{\text{scr}} = 4.06$] and after [$M_{\text{con}} = 19.57, SD_{\text{con}} = 2.64; M_{\text{rub}} = 19.43, SD_{\text{rub}} = 3.16; M_{\text{scr}} = 19.49, SD_{\text{scr}} = 3.43$]. That is, all groups significantly increased their SSRL, with no differences between them.

RQ2: What are the effects of rubrics and scripts on self-efficacy?

No significant differences were found between the three groups across the measured time points [$F(2, 111) = 2.15, p = 0.12, \eta^2 = 0.04$]. However, we found a significant difference between the control group and the rubric group ($p = .022$) before the intervention [$F(2, 121) = 4.07, p = .019, \eta^2 = .063; M_{\text{con}} = 12.93, M_{\text{rub}} = 11.27, M_{\text{scr}} = 12.59$]. Crucially, after the intervention the differences were not significant [$F(2, 117) = .219, p = .80, \eta^2 = .004; M_{\text{con}} = 13.27, M_{\text{rub}} = 12.75, M_{\text{scr}} = 12.88$], largely because of the higher increase in the self-efficacy of the rubric condition. Moreover, the occasion was significant only for the rubric group [$F(1, 111) = 6.66, p = 0.01, \eta^2 = 0.057, M_{\text{pre}} = 11.27, M_{\text{post}} = 12.75$]. In summary, the control group and the script group maintained their levels of self-efficacy and in the rubric group it increased significantly with the intervention.

The results of the first three RQ can be seen in Figure 1.

<Insert Figure 1 around here>

RQ3: What are the effects of rubrics and scripts on academic performance?

The conditions were compared in three different assessment practices. First, in the three task assignments we found significant differences (see Table 1) [$F(2, 125) = 7.75, p = .001, \eta^2 =$

.11; $M_{\text{con}} = 8.52$, $M_{\text{rub}} = 8.22$, $M_{\text{scr}} = 7.48$]. In each of the tasks, the control group obtained the highest scores, followed by the rubric group and, with significantly lower scores, the script group. Second, in the partial examination we also found significant differences [$F(2, 125) = 7.70$, $p = .001$, $\eta^2 = .11$; $M_{\text{con}} = 5.47$, $M_{\text{rub}} = 3.68$, $M_{\text{scr}} = 4.39$]. And third, in the final examination, we did not find significant differences at the 95% confidence level, though we did find them at the 90% level [$F(2, 118) = 2.60$, $p = .08$, $\eta^2 = .04$; $M_{\text{con}} = 5.82$, $M_{\text{rub}} = 4.82$, $M_{\text{scr}} = 4.98$].

<Insert Table 1 around here>

Given that the control group had a significantly higher starting level (grade in the previous academic year. See the opening of *Results*), it is necessary to apply some techniques that can eliminate this initial bias. In order to do so, we applied a difference-in-differences analysis of the measures of the grades in the multiple time periods (see *Data analysis*), and the segmentation of the sample into quartiles according to the grade obtained in the similar course of the previous academic year and also according to the grade of the academic record (in order to check whether the possible impact of the instrument was significantly different depending on the rank of students' performance). The results showed the following significant results (see Supplementary Material C):

In temporary change 1, previous academic year's course grade vs final course grade of this research, for the total sample, the scripts showed a negative significant effect ($\beta = -1.205$, $p = .013$). By segmenting the sample, a positive significant effect was identified for students with a very high academic record grade who employed the rubrics ($\beta = 2.366$, $p = .068$).

In temporary change 2, grade of midterm examination vs final grade examination in the course of this research, when segmenting the sample, a positive significant effect was identified

in students with a very high academic record grade who employed rubrics ($\beta = 2.26, p = .06$). In addition, positive significant effects were found for students who employed rubrics with an intermediate grade ($\beta = 3.013, p = .01$) or a high grade ($\beta = 3.430, p = .000$) in the similar course from the previous academic year; likewise for students in the scripts group who had a grade in a similar course from the previous year that was intermediate ($\beta = 2.316, p = .022$) or high ($\beta = 2.498, p = .012$).

In temporary change 3, grade of the first group assignment vs grade of third group assignment in the course of this research, significant results were obtained for the total sample in the use of rubrics ($\beta = 1.993, p = .001$). In addition, positive significant effects were found for students who employed rubrics with a high academic record grade ($\beta = 1.971, p = .028$) or with a grade in a similar course from the previous year that was very high ($\beta = 2.143, p = .01$).

In short, these analyses showed that the use of rubrics, in particular, generated positive effects, especially for those students with a previous very high academic record grade.

RQ4: What are the students' perceptions of, and how do they report, employing the rubrics/scripts?

The results of the quantitative questionnaire (Table 2) show that the students generally expressed a moderate degree of agreement regarding what the use of rubrics/scripts brought them. This highlights that the three highest rated items are those related to understanding what is asked of them (items 1, 2 and 5) and the lowest rated items related to grading and learning (items 3, 4, 7 and 8).

In addition, compared to the rubric group, the script group reported with significantly higher levels that the use of the tool helps to understand teacher expectations [$F(1, 83) = 5.487, p = .022, \eta^2 = .062; M_{\text{rub}} = 2.62, M_{\text{scr}} = 3.14$], helps to plan and tailor the task to what is

requested [$F(1, 83) = 7.864, p = .006, \eta^2 = .087; M_{\text{rub}} = 2.45, M_{\text{scr}} = 3.09$], and to optimize time and be more productive [$F(1, 83) = 6.632, p = .012, \eta^2 = .074; M_{\text{rub}} = 1.93, M_{\text{scr}} = 2.56$].

<Insert Table 2 around here>

Table 3 shows the frequencies of the grouped categories derived from the open-ended questions. See Supplementary Material D for a detailed exploration of the five categories, their 25 subcategories and frequencies, and Supplementary Material E for the most representative verbatim of these. The main benefit, stated by 82.2% of the participants of the rubric group and 95.7% of the script group, was related to self-regulation and some of its phases. Another benefit, such as obtaining a higher grade, was only stated by students in the rubric group. The main disadvantage was the limitation of the personal way of doing the assignment.

<Insert Table 3 around here>

The students reported a moderate use of the rubrics/scripts, in general, in the three items associated with the three phases of self-regulation, and also in the use of the rubrics/scripts for the completion of the three assignments (see Table 4).

<Insert Table 4 around here>

Discussion

Our aim was to explore the effects of rubrics and scripts on students' self-regulation (RQ1), self-efficacy (RQ2), performance (RQ3), and how students perceived and used (RQ4) these two formative assessment tools. Although previous research on the use of rubrics and scripts showed mixed results, our findings are certainly positive.

There are three crucial aspects to consider regarding the discussion. To the best of the authors' knowledge, only three previous investigations have directly explored the effects of rubrics and scripts in the same study. However, their designs presented some characteristics that

called for further investigation. First, Panadero et al. (2012) was implemented in a laboratory context, Panadero et al. (2013) was a quasi-experiment with real-classroom data with a limited participant enrolment (20 in the experimental rubric and script groups and 29 in the control), and Panadero et al. (2014) was also a quasi-experiment with real-classroom data, involving a rubric and a script group, without a control. Second, importantly in those three studies the tasks performed by the participants were individual while here the tasks were performed in groups. Third, the tasks in the research by Panadero et al. (2012, 2013, 2014) were, respectively, a landscape analysis, design multimedia material and summarised texts, and designed conceptual maps. In our research, the three tasks were complex econometric problems the solutions to which were not simple.

Self-regulated learning

With regard to RQ1, the results showed that the overall levels of SRL measured via the MSLQ questionnaire were similar in the pre-post intervention in the rubrics and scripts group and decreased in the control group. Given the higher achievement levels of the control condition, the fact that the two conditions using the tools outperformed the control indicates that these increased the participants' use of strategies. In the specific self-regulation questionnaire, all groups significantly increased their SSRL, with no differences between them. Our interpretation is that in such specific SRL strategies that were closer to the task than the MSLQ measurement, the control with its higher achievement levels should have outperformed the tool conditions. Therefore, our overall interpretation is that both tools had positive effects, even if they did not outperform the control in the two measures.

In the three studies by Panadero et al. (2012, 2013, 2014), they also obtained mixed results in their use of different approaches to measure SRL. Using the EMSR-Q questionnaire, in

two of the three studies (2013, 2014) the script group benefited in their Learning SRL and the rubric group benefited in their Avoidance SRL, whereas in the 2012 study no differences were found. Interestingly, this study (2012) was the only one to use think-aloud protocols, with the rubric group scoring higher than the control group – although not significantly – and the script group performing significantly more self-regulatory actions than the rubric group. This aspect is fully in line with the meta-analysis of Panadero et al. (2017), concluding that when SRL was measured using qualitative data instead of questionnaires, the effect of self-assessment was greater.

In addition to this, the tasks in which rubrics and scripts were employed in this study were performed in groups. In comparison, Panadero et al. (2012, 2013, 2014) employed individual tasks. Collaborative learning can be an effective approach to learning, but its success depends on the task complexity and prior knowledge of the group members (Kirschner et al., 2018). When undertaking high-complexity tasks, group learning is often more effective than individual learning because peers can use their primary social skills, such as communication and coordination, to share their working memory resources. However, for low-complexity tasks, collaboration can be irrelevant and even detrimental, as it can impose additional cognitive load. In such cases, individual learning may be more effective as it allows learners to focus on the task at hand without the distraction of coordinating with others (Kirschner et al., 2011). Therefore, for the complex problem-solving tasks in this study, collaborative task performance may have been a positive aspect explaining our findings on SRL.

Self-efficacy

With regard to RQ2, self-efficacy, the control group and the script group maintained their levels of self-efficacy and in the rubric group it increased significantly after the intervention.

First, Panadero et al. (2012) showed positive effects with the triple interaction of tool/feedback/occasion, especially in the use of the rubric. In contrast, in the other two studies (Panadero, 2013, 2014), where no feedback was provided, no significant effects were found. The present research had similar results, in which three similar activities were performed, and group feedback was provided in class, in addition to the grades. These teaching actions seem to have no positive results compared to providing feedback (van Dinther et al., 2011; Panadero et al., 2012).

Performance

In RQ3, the control had a significantly higher achievement level in the previous course, which was the introduction for the course in which we conducted our intervention. For that reason, we ran two types of data analysis. In the first, the ANOVA family analysis, we did not counter differences in those baselines, and the control group also performed significantly better.

Second, to account for the control condition baseline superiority, we performed a difference-in-differences analysis over three time periods and the segmentation of the sample into quartiles. Mainly, positive significant effects were identified for students who used rubrics, particularly for those with a very high academic grade record, and so too for students in the scripts group who had an intermediate or high grade in the previous year's similar course. Also taking into account the results in SRL, it could be considered that students starting from a higher academic performance are more able to self-regulate and better employ self-assessment tools.

Dunning et al. (2003) suggest that low-performing students may not be aware of their inadequate performance, leading to a lack of recognition that impedes academic success and hinders self-regulated learning. This idea is supported by García-Pérez et al. (2021), who found that all participants with low academic performance had limited knowledge of learning strategies.

Therefore, since SRL plays a crucial role in the academic performance of students with different

levels of prior achievement (Hirt et al., 2021), teachers can implement strategies to enhance SRL such as those suggested in this discussion and in future lines of research.

In addition, more positive effects were found for students who used rubrics. The explanation for this may lie in the fact that the rubric is closer to what the task/product should look like (Brookhart, 2018) and is an easier instrument to employ than scripts because scripts require the activation of more strategies (Kollar et al., 2007).

Compared to previous similar research, in the results of Panadero et al. (2014) the rubric group outperformed the script group, although there was no control group. Panadero et al. (2012) did find significant differences, with the rubric and script groups outperforming the control group. Panadero et al. (2013) found no difference between the three groups (rubric, script, and control), although there was the limitation that performance was measured only through the course final grade. In addition, Panadero et al.'s (2013) study and our research had many similarities: the tasks were very well defined and structured so the control group could also benefit from this; and the tasks were also performed in groups. However, the difficulty of the tasks in this study is high and, as we indicated in the discussion of self-regulated learning, task complexity plays an important role. Learning from high-complexity tasks in groups may be better than individual learning, while low-complexity tasks are better learned individually because collaboration can be detrimental to learning (Kirschner et al., 2011). Compared to previous similar studies, such as Panadero et al. (2013), this may explain the positive results observed in this study.

Perception and use of rubrics and scripts

Students reported a medium degree of agreement with the questionnaire statements about the use of rubrics or scripts. In comparison, scripts received more positive ratings than rubrics,

especially in relation to understanding the teacher's expectations, and planning and tailoring the assignment. This result is in line with the coding of their opinions in the open-ended questions. Here we would like to draw attention to three aspects: the usability of each tool; contextual aspects of the instructional setting; and the nature of the tasks performed by the students. First, it seems that students in the script group might have a false perception of the tool as the rubrics showed higher positive effects in this research. This may be because scripts are considered a simpler tool (Kollar et al. 2007), although in practice scripts are more cognitively demanding (Panadero et al., 2013). Only in the study by Panadero et al. (2014) students' perceptions of the two tools were compared. The use of rubrics was perceived as more helpful in the three SRL phases of the task in Panadero et al. (2014). However, in this study we did not find significant differences. Second, as discussed above, aspects such as the context of the task, working in a group, or the provision of feedback are reasons that point to why students did not really need to employ these tools and thus consider them useful. Third, the nature of the task may be an essential aspect to consider in terms of the effectiveness of these tools. In the studies by Panadero et al. (2012, 2013, 2014), the tasks were writing summaries and designing conceptual maps, and the rubrics and scripts precisely convey the different parts that are required in the task and what the quality should be (see Appendix A of Panadero et al. (2014)). However, problem-solving rubrics can only help to attend to the details of the problem-solving process (Hull et al., 2013).

Finally, interestingly, in the coding of the qualitative data many more students in the rubric group than in the script group considered that it would have been useful to create the tool with the teacher. This trend is also seen in the quantitative data. One possible reason is that, when comparing the perception of both groups, students in the script group perceived this tool to

be more useful, and is also easier to use (Atkinson et al., 2003) than the rubric. This finding is in line with the positive effects shown in the self-regulation and academic performance of students who co-created and used three rubrics rather than simply employing them (Fraile et al., 2017).

Limitations

The study's reliance on self-reported data is one of its major limitations. According to earlier studies, measuring self-regulation is more valid when it is compared to multiple forms of data, namely process data (e.g. thinking-aloud protocols). In addition, it should be considered that previous similar studies only obtained significant results when measuring SRL through thinking-aloud protocols. A possible additional limitation is that students in different groups may have exchanged rubrics or scripts. However, we took some measures to try to prevent this situation (see *Participants* and *Procedure*).

Future lines of research

First, in order to improve the efficiency and effectiveness of collaborative learning environments and the use of formative assessment tools, it is essential to investigate how student groups work and to consider the impact of cognitive load. As noted by Kirschner et al. (2018), despite its valuable insights into the mechanisms of collaborative learning, cognitive load theory is often overlooked in the design and research of such settings. First and foremost is the importance of transactive activities (Popov et al., 2017). This involves facilitating the acquisition of individual and group-specific knowledge as well as shared general knowledge. Second, accordingly, the features of each constituent factor play an important role in determining the dynamics of a collaborative learning scenario. These three factors (Kirschner et al., 2018), namely task characteristics and guidance, learner characteristics (domain-specific expertise and

collaborative skills), and team characteristics (size, roles, composition, and prior team experience collaborating on similar tasks), shape the unique aspects of each collaborative learning situation.

Secondly, there are numerous studies on “collaborative or external scripts”, which focus on the social interactions that take place during collaborative learning, but do not contain any information about the topic to be taught (Laru et al., 2012), unlike the (internal) scripts used in this study. This means that collaborative or external scripts are used to improve the way the group works together, but not to support the task, which is the aim of internal scripts. In fact, a meta-analysis of external scripts showed a large positive effect on collaborative skills, but a small positive effect on domain-specific knowledge (Vogel et al., 2017). Therefore, future research could explore the combined use of internal scripts and collaboration or external scripts.

Third, although rubrics have been used effectively for written tasks, there is little research on how applicable they are to problem-solving tasks, which require a different approach to completing the task.

Fourth, co-creation of the instruments between the teacher and the students seems to be a promising action to enhance SRL and performance rather than just directly providing the instruments themselves, at least in rubrics (e.g. Fraile et al., 2017).

Fifth, given the different purposes of rubrics and scripts, future research should empirically explore their complementary rather than comparative use. In this vein, the questionnaires could provide useful data, but since they are presented in terms of rubric versus script, valuable information about each is overlooked.

Sixth, the provision of feedback in these activities needs to be investigated, either by the teacher – as their feedback seems to be positive in rubric and script studies (Panadero et al., 2012) – or through self- and peer assessment activities. Furthermore, according to the systematic

review by To et al. (2021), the use of exemplars, only in combination with rubrics, peer, and self-assessment, enhances student performance.

Seventh, in relation to SRL, other tools and approaches could be considered, one of which is the use of video models. These are a more dynamic tool that provides explicitly key criteria and effective tactics for solving problems and picking new assignments (Allal 2020).

Finally, to make research designs in assessment studies more robust, it is important to incorporate elements such as the presence of a control group, experimental design, conducting the study in a classroom setting rather than a laboratory setting, ensuring a sufficiently large sample size, conducting longitudinal studies that examine the effects over a longer period of time, controlling for previous measures such as self-regulated learning profiles and previous performance, and using more objective measures such as thinking-aloud protocols, in addition to self-reported measures.

Conclusions

Considering the results as a whole, the use of rubrics showed strong positive results and the use of scripts slightly positive benefits on both academic performance and self-efficacy. Interestingly, students in the rubric group who started from a very high academic level were the ones who benefited greatly from the use of this tool. Slightly differently, significant results were obtained for students who used scripts and had an intermediate or high starting level. In addition, the study found that only the rubric was an effective tool for promoting self-efficacy. Future research is needed with robust research designs that consider the context of the collaborative tasks, the difficulty of the task, students' prior performance, and teachers' and students' experience in using the tools. The positive effects could be maximised with other strategies, such

as the provision of feedback, self- and peer assessment activities, the use of exemplars, video models or the co-creation of these formative assessment instruments together with the teacher.

Acknowledgements

We thank Ernesto Panadero for his great help in the design and writing of this article. We are grateful to Heidi Andrade for the feedback in the SIG1 & SIG4 Conference 2022. Finally, we would like to thank Javier Fernández-Ruiz and José M. Sánchez Galán for their collaboration with data collection and processing.

References

- Allal, L. (2020). Assessment and the co-regulation of learning in the classroom. *Assessment in Education: Principles, Policy & Practice*, 27(4), 332–349.
<https://doi.org/10.1080/0969594X.2019.1609411>
- Alonso-Tapia, J., & Panadero, E. (2010). Effect of self-assessment scripts on self-regulation and learning. *Infancia y Aprendizaje*, 33(3), 385–397.
<https://doi.org/10.1174/021037010792215145>
- Alonso-Tapia, J., Panadero, E., & Ruiz, M. A. (2014). Development and validity of the Emotion and Motivation Self-regulation Questionnaire (EMSR-Q). *The Spanish Journal of Psychology*, 17, e55. <https://doi.org/10.1017/sjp.2014.41>
- Andrade, H. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, 4, 87. <https://doi.org/10.3389/feduc.2019.00087>
- Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice*, 48(1), 12–19.
<https://doi.org/10.1080/00405840802577544>

- Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology, 95*(4), 774–783. <https://doi.org/10.1037/0022-0663.95.4.774>
- Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology: An International Review, 54*(2), 199–231. <https://doi.org/10.1111/j.1464-0597.2005.00205.x>
- Brookhart, S. M. (2018). Appropriate Criteria: Key to Effective Rubrics. *Frontiers in Education*. <https://doi.org/https://doi.org/10.3389/educ.2018.00022>
- Callaway, B., & Sant’Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics, 225*(2), 200–230. <https://doi.org/10.1016/j.jeconom.2020.12.001>
- Dawson, P. (2017). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education, 42*(3), 347–360. <https://doi.org/10.1080/02602938.2015.1111294>
- Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*(3), 83–87. <https://doi.org/10.1111/1467-8721.01235>
- Fraile, J., Panadero, E., & Pardo, R. (2017). Co-creating rubrics: The effects on self-regulated learning, self-efficacy and performance of establishing assessment criteria with students. *Studies in Educational Evaluation, 53*, 69–76. <https://doi.org/10.1016/j.stueduc.2017.03.003>

- García-Pérez, D., Fraile, J., & Panadero, E. (2021). Learning strategies and self-regulation in context: how higher education students approach different courses, assessments, and challenges. *European Journal of Psychology of Education, 36*(2), 533–550.
<https://doi.org/10.1007/s10212-020-00488-z>
- Hirt, C. N., Karlen, Y., Merki, K. M., & Suter, F. (2021). What makes high achievers different from low achievers? Self-regulated learners in the context of a high-stakes academic long-term task. *Learning and Individual Differences, 92*, 102085.
<https://doi.org/10.1016/j.lindif.2021.102085>
- Hull, M. M., Kuo, E., Gupta, A., & Elby, A. (2013). Problem-solving rubrics revisited: Attending to the blending of informal conceptual and formal mathematical reasoning. *Physical Review Special Topics-Physics Education Research, 9*(1), 010105.
<https://doi.org/10.1103/PhysRevSTPER.9.010105>
- Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assessment & Evaluation in Higher Education, 39*(7), 840–852.
<https://doi.org/10.1080/02602938.2013.875117>
- Kirschner, F., Paas, F., & Kirschner, P. A. (2011). Task complexity as a driver for collaborative learning efficiency: The collective working-memory effect. *Applied Cognitive Psychology, 25*(4), 615-624. <https://doi.org/10.1002/acp.1730>
- Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano R, J. (2018). From cognitive load theory to collaborative cognitive load theory. *International Journal of Computer-Supported Collaborative Learning, 13*, 213–233. <https://doi.org/10.1007/s11412-018-9277-y>

- 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.
<https://doi.org/10.1016/j.learninstruc.2007.09.021>
- Laru, J., Järvelä, S., & Clariana, R. B. (2012). Supporting collaborative inquiry during a biology field trip with mobile peer-to-peer tools for learning: a case study with K-12 learners. *Interactive Learning Environments, 20*(2), 103–117.
<https://doi.org/10.1080/10494821003771350>
- 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. <https://doi.org/10.1016/j.lindif.2012.04.007>
- Panadero, E., Alonso-Tapia, J., & Huertas, J. A. (2014). Rubrics vs. self-assessment scripts: effects on first year university students' self-regulation and performance / Rúbricas y guiones de autoevaluación: efectos sobre la autorregulación y el rendimiento de estudiantes universitarios de primer año. *Infancia y Aprendizaje, 37*(1), 149–183.
<https://doi.org/10.1080/02103702.2014.881655>
- Panadero, E., Alonso-Tapia, J., & Reche, E. (2013). Rubrics vs. self-assessment scripts effect on self-regulation, performance and self-efficacy in pre-service teachers. *Studies in Educational Evaluation, 39*(3), 125–132. <https://doi.org/10.1016/j.stueduc.2013.04.001>
- Panadero, E., Brown, G. T. L., & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review, 28*(4), 803–830. <https://doi.org/10.1007/s10648-015-9350-2>

- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129–144.
<https://doi.org/10.1016/j.edurev.2013.01.002>
- Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, 22, 74–98.
<https://doi.org/10.1016/j.edurev.2017.08.004>
- Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice*, 21(2), 133-148. <https://doi.org/10.1080/0969594X.2013.877872>
- Pintrich, P. R., Smith, D. A., García, T., & McKeachi, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire*. The Regents of the University of Michigan.
- Popov, V., van Leeuwen, A., & Buis, S. C. A. (2017). Are you with me or not? Temporal synchronicity and transactivity during CSCL. *Journal of Computer Assisted Learning*, 33(5), 424–442. <https://doi.org/10.1111/jcal.12185>.
- Reynolds-Keefer, L. (2010). Rubric-referenced assessment in teacher preparation: An opportunity to learn by using. *Practical Assessment, Research & Evaluation*, 15(8).
<https://doi.org/10.7275/PSK5-MF68>
- Roth, A., Ogrin, S., & Schmitz, B. (2016). Assessing self-regulated learning in higher education: a systematic literature review of self-report instruments. *Educational Assessment, Evaluation and Accountability*, 28(3), 225–250. <https://doi.org/10.1007/s11092-015-9229-2>

- 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. <https://doi.org/10.1348/000709906X106147>
- To, J., Panadero, E., & Carless, D. (2021). A systematic review of the educational uses and effects of exemplars. *Assessment & Evaluation in Higher Education*, 1–16. <https://doi.org/10.1080/02602938.2021.2011134>
- van Dinther, M., Dochy, F., & Segers, M. (2011). Factors affecting students' self-efficacy in higher education. *Educational Research Review*, 6(2), 95–108. <https://doi.org/10.1016/j.edurev.2010.10.003>
- Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review*, 29(3), 477–511. <https://doi.org/10.1007/s10648-016-9361-7>
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Yan, Z., Lao, H., Panadero, E., Fernández-Castilla, B., Yang, L., & Yang, M. (2022). Effects of self-assessment and peer-assessment interventions on academic performance: A meta-analysis. *Educational Research Review*, 37, 100484. <https://doi.org/10.1016/j.edurev.2022.100484>
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 65–70. https://doi.org/10.1207/s15430421tip4102_2

Figure 1

Plots of Self-regulation and Self-efficacy Data

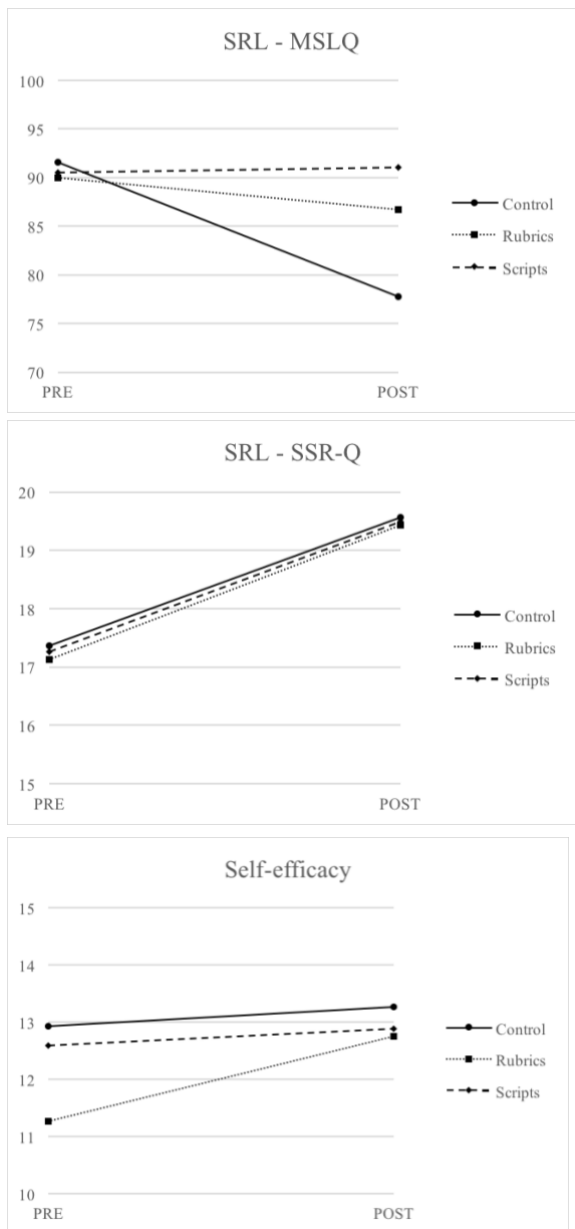


Table 1*Course Grades*

	Control	Rubric	Script		
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>p</i>	η^2
Assignment 1	8.76(1.22)	8.06(1.74)	7.84(1.27)	.008	.072
Assignment 2	8.05(1.51)	7.70(1.97)	7.29(1.95)	.171	.028
Assignment 3	8.70(1.89)	8.79(1.38)	6.74(2.46)	.000	.197
Assignments mean	8.52(1.07)	8.22(1.38)	7.48(1.25)	.001	.11
Partial examination	5.47(2.20)	3.68(2.10)	4.39(1.92)	.001	.11
Final examination	5.82(1.96)	4.82(2.50)	4.98(1.67)	.078	.042

Table 2*Quantitative Questionnaire on Students' Perceptions of Rubrics/Scripts*

Item	Rubrics		Scripts		<i>p</i>	η^2
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
1. Using rubrics/self-assessment scripts helps me understand the teacher's expectations	2.62	1.29	3.14	0.68	.022	.062
2. Using rubrics/scripts helps me to plan and tailor my work to what is asked for	2.45	1.29	3.09	0.75	.006	.087
3. With rubrics/scripts I get a higher grade than if I did not have them	2.19	1.25	2.56	1.14	.161	.024
4. Using rubrics/scripts helps me to optimize the time I spend at work and to be more productive	1.93	1.28	2.56	0.96	.012	.074
5. Rubrics/scripts help me self- assess my work	2.76	1.25	3.02	0.94	.277	.014
6. Using rubrics/scripts makes my work of higher quality	2.45	1.21	2.70	1.06	.323	.012
7. Using rubrics/scripts makes me less nervous about the grade I will get	1.83	1.27	1.81	1.31	.945	.000

8. Having rubrics/scripts makes me learn more than if I did not have them	2.20	1.27	2.44	1.03	.330	.012
9. My overall rating of the use of rubrics/scripts is positive	2.66	1.22	2.77	0.97	.651	.003
10. It would be better to design the rubrics/scripts with the teacher vs having the teacher provide them directly	2.00	1.41	1.84	1.27	.578	.004

Note: 5 five-point Likert scale (“completely disagree”–“completely agree”) from 0 to 4.

“Rubrics/scripts” appears in the items, although the questionnaire completed by the students only showed the instrument used by that group.

Table 3*Qualitative Questionnaire on Students' Perceptions of Rubrics/Scripts*

Categories	Rubrics		Scripts	
	<i>n</i>	%	<i>n</i>	%
Benefits	45	24.19	46	24.73
Disadvantages	27	14.52	23	12.37
Use of the tool	43	23.12	38	20.43
Anxiety/stress	35	18.82	42	22.58
Co-create with the teacher	36	19.35	37	19.89
Total	186	100	186	100

Table 4*Questionnaire for Students' Use of Rubrics/Scripts*

Item	Rubrics		Scripts		<i>p</i>	η^2
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
In general, how much have you used self-assessment rubrics/scripts?	2.31	1.14	2.47	0.88	.48	.006
In general, I have used the rubrics/scripts to plan the tasks I had to do in the assignment <i>[First phase of SRL – Forethought, planning]</i>	2.43	1.11	2.51	1.08	.73	.001
In general, I have used the rubrics/scripts to guide me while I was doing the task and to check if I was on the right/wrong path <i>[Second phase of SRL – Monitoring]</i>	2.21	1.14	2.37	1.18	.53	.005
In general, I have used the rubrics/scripts to evaluate what I had accomplished and what I had not in the task once I had completed it <i>[Third phase of SRL – Self-evaluating]</i>	2.67	1.16	2.74	1.18	.76	.001

In assignment 1, how much did you use the rubrics/scripts?	2.73	1.15	2.67	1.11	.84	.001
In assignment 2, how much did you use the rubrics/scripts?	2.68	1.11	2.70	1.06	.95	.000
In assignment 3, how much did you use the rubrics/scripts?	2.85	1.13	2.84	1.09	.95	.000

Note. Five-point Likert scale (“I have not used them”–“I have used them a lot”) from 0 to 4. “Rubrics/scripts” appears in the items, although the questionnaire completed by the students only showed the instrument used by that group.

Supplementary Material A

Self-assessment script of mandatory assignment 1: “Steps in regression analysis”

	YES	NO
1. Am I clear about the task to be performed in each section of this activity?		
Regarding the specification stage		
2. To specify the model, have I identified the variables that explain the endogenous variable with causality?		
3. Have I checked whether these variables are in the work file and whether their measurement (codes for the values of that variable) is the correct one to include in an econometric model?		
4. Have I generated in Eviews the new or transformed variables I need?		
Regarding the estimation and validation stages		
5. Have I performed and interpreted the correlation analysis prior to the estimation of an econometric model?		
6. Have I performed the interpretation of the estimation output: sign, individual significance, joint goodness of fit and residual analysis?		
7. Have I interpreted whether the results obtained in the estimation output are correct or whether they need to be improved through a new re-specification of the econometric model (validation stage)?		
8. Have I re-specified the model as many times as necessary to obtain the best possible fit?		
Regarding the utilization stage		
9. Have I interpreted the sign and amount of the estimated parameter?		
10. And have I constructed and interpreted the confidence interval of the parameter?		

11. From the results obtained, have I identified the order of influence of the explanatory variables on the endogenous variable?		
12. Have I performed and interpreted one or more tests on individual parameter restrictions, for null hypotheses other than equality to zero?		
13. Have I performed and interpreted one or more tests on parameter set restrictions, for null hypotheses other than equality to zero?		
14. Have I made one or more predictions using the results obtained in the econometric model?		
15. And have I calculated the confidence interval of the prediction(s)?		
16. Have I compared the results obtained with those of an alternative estimate (e.g. those of a specific Autonomous Community)?		

Supplementary Material B

Rubric of mandatory assignment 1: “Steps in regression analysis”

	EXCELLENT	SATISFACTORY	REQUIRES IMPROVEMENT	POOR
SPECIFICATION STAGE				
Selection of variables	All explanatory “causal” variables were selected and the necessary transformations were performed on the data.	Most of the “causal” explanatory variables have been selected and the necessary transformations have been performed on the data.	Most of the explanatory “causal” variables have been selected but it is not known how to perform the necessary transformations on the data.	Failure to correctly identify the explanatory “causal” variables and to perform the necessary transformations on the data
ESTIMATION AND VALIDATION STAGES				
Correlation analysis	Correlation analysis has been performed in Eviews and the results have been interpreted in such a way as to improve	Correlation analysis has been performed in Eviews and the results have been interpreted.	Correlation analysis has been performed in Eviews but the results have not been correctly interpreted.	Correlation analysis has not been performed

	the initial specification of the model.			
Results obtained in model estimation and validation	The estimation of the model has been carried out and all the results obtained in the validation have been interpreted: signs, individual significance, joint goodness of fit and analysis of residuals.	The estimation of the model has been carried out and most of the results obtained in the validation have been interpreted: signs, individual significance, joint goodness of fit and analysis of residuals.	The estimation of the model has been carried out and some but not all of the results obtained have been interpreted: signs, individual significance, joint goodness of fit and residual analysis.	Model estimation has been performed but the results have not been interpreted.
Respecification of the model	The validation results have been used to respecify the model as many times as necessary to obtain the best possible fit.	Validation results have been used to respecify the model and obtain a good fit.	Few validation results have been obtained, so the initial specification has been practically unchanged.	No changes to the initial specification have been made
USE PHASE				

Interpretation of the estimated parameter	<p>The sign and amount of the estimated parameter have been interpreted. Its CI has been constructed and interpreted.</p>	<p>The sign and amount of the estimated parameter have been interpreted. Its CI has been constructed but not interpreted.</p>	<p>The sign and amount of the estimated parameter have been interpreted, but its CI has not been constructed.</p>	<p>Only the sign of the estimated parameter has been interpreted.</p>
Ranking of explanatory variables	<p>The order of influence of all the explanatory variables on the endogenous variable has been identified.</p>	<p>The order of influence of most of the explanatory variables on the endogenous variable has been identified.</p>	<p>The order of influence of some explanatory variables on the endogenous variable has been identified, but not all of them.</p>	<p>The order of influence of the explanatory variables on the endogenous variable cannot be identified.</p>
Parameter restriction	<p>Several tests have been carried out on individual and parameter set restrictions, for null hypotheses other than equality at zero.</p>	<p>A single-parameter constraint and a set of parameters' constraint have been tested for null hypotheses other than equality at zero.</p>	<p>Contrasts have been performed on individual and parameter set restrictions, for null hypotheses of equality to zero.</p>	<p>No contrasts on parameter restrictions have been performed.</p>

Prediction	Several point predictions have been made and the CI of each one has been calculated.	A point prediction has been made and its CI has been calculated.	A point prediction has been made but its CI has not been calculated.	It is not known how to use the model to make a prediction.
Comparison of models	Alternative estimates have been made for different data samples and the results obtained in their output have been compared. Contrasts on parameter restrictions affecting the alternative estimations have also been carried out.	Alternative estimations have been performed for different data samples and the results obtained in the output have been compared.	Alternative estimates have been made for different data samples, but the results obtained have not been compared.	Alternative estimates have not been made for different data samples.

Supplementary Material C

Difference-in-differences with multiple time periods

Temp	Segme	Quartile	Rubric group vs. control group				Script group vs. control group			
			B	β	t	<i>p</i>	B	β	t	<i>p</i>
			(SE)				(SE)			
1	1	All	-0.30	-0.08	-0.54	0.59	-1.21	-0.34	-2.50	0.01*
			(0.56)				(0.48)			**
		Low	-1.22	-0.42	-0.99	0.33	-1.94	-0.71	-1.68	0.11
			(1.22)				(1.15)			
		Intermedi	0.17	0.05	0.18	0.86	-1.19	-0.38	-1.29	0.21
	ate	(0.96)				(0.92)				
	High	-0.48	-0.11	-0.40	0.69	-1.27	-0.31	-1.32	0.19	
		(1.19)				(0.96)				
	Very high	2.37	0.48	1.91	0.07*	-0.69	-0.24)	-0.89	0.38	
		(1.24)				(0.78)				
2	1	All	0.60	0.12	0.91	0.36	-0.13	-0.03	-0.23	0.82
			(0.66)				(0.54)			
		Low	-0.60	-0.17	-0.41	0.68	-0.89	-0.30	-0.67	0.51
			(1.45)				(1.33)			
		Intermedi	1.04	0.23	0.77	0.45	-0.10	-0.04	-0.12	0.90
	ate	(1.35)				(0.83)				
	High	0.74	0.18	0.79	0.43	-0.30	-0.07	-0.33	0.74	
		(0.93)				(0.91)				
	Very high	2.26	0.50	1.97	0.06*	0.37	0.12	0.46	0.65	

			(1.15)			(0.80)				
2		All	0.60	0.12	0.91	0.36	-0.13	-0.03	-0.23	0.82
			(0.66)				(0.54)			
		Low	2.44	0.61	1.13	0.27	2.19	0.76	1.54	0.14
			(2.16)				(1.43)			
		Intermedi ate	3.01	0.85	2.78	0.01*	2.32	0.92	2.41	0.02*
			(1.09)			**	(0.96)			*
		High	3.43	0.73	4.87	<0.01	2.50	0.44	2.63	0.01*
			(0.71)			***	(0.95)			**
		Very high	1.30	0.29	1.47	0.15	1.14	0.20	1.16	0.26
			(0.89)				(0.99)			
3	1	All	1.99	0.43	3.33	<0.01	-0.62	-0.15	-1.19	0.24
			(0.60)			***	(0.52)			
		Low	3.22	0.58	1.62	0.11	-0.63	-0.16	-0.37	0.71
			(1.99)				(1.69)			
		Intermedi ate	1.25	0.28	1.02	0.31	-0.96	-0.24	-0.88	0.39
			(1.22)				(1.13)			
		High	1.97	0.58	2.29	0.03*	-0.35	-0.08	-0.34	0.74
			(0.86)			*	(1.03)			
		Very high	-0.98	-0.31	-1.06	0.30	-0.66	-0.17	-0.77	0.45
			(0.93)				(0.85)			
	2	All	1.99	0.43	3.33	<0.01	-0.62	-0.15	-1.19	0.24
			(0.60)			***	(0.52)			
		Low	2.10	0.40	0.81	0.43	-1.56	-0.42	-0.84	0.41
			(2.61)				(1.86)			
		Intermedi ate	1.77	0.36	1.06	0.30	-2.35	-0.58	-1.42	0.17
			(1.67)				(1.66)			

High	1.47	0.33	1.45	0.16	-0.74	-0.12	-0.77	0.45
	(1.02)				(0.97)			
Very high	2.14	0.65	2.72	0.01*	0.96	0.21	0.97	0.34
	(0.79)			**	(0.99)			

Note. Temporary change: (1) the final grade of the subject from the previous academic year to the final grade of the course of this research; (2) from the midterm to the final exam in the course of this research; and (3) from the first group assignment to the third group assignment in the course of this research. Segment: students were segmented by quartiles according to the (1) academic record grade and (2) to the grade obtained in the similar subject in the previous academic year (low–intermediate–high–very high). For time period 1, there is only one segment since the other involves a time period 1 variable (grade in similar subject in the previous academic year). ***, **, and * indicate significance at 1%, 5% and 10% levels, respectively.

Supplementary Material D

Detailed exploration of the five categories and their 25 subcategories and frequencies

Categories	Rubrics		Scripts	
	<i>n</i>	%	<i>n</i>	%
Benefits	45	100	46	100
Useful for self-regulation (being aware of what needs to be done in the task; being able to organize oneself; reflecting on whether one is doing well; identifying failures; checking at the end; useful when one does not know how to proceed, etc.)	37	82.2	44	95.7
Higher grade	5	11.1	0	0
Improving learning (checking what has been learned; understanding what is being done)	2	4.4	2	4.3
Rely on trustworthy material provided by the teacher.	1	2.2	0	0
Disadvantages	27	100	23	100
Limiting the personal way of doing the task (if you cannot develop further a part of the practice you master better; does not allow you to go deeper into something you consider important; tools schematize a lot)	9	33.3	4	17.4
Limitation of the rubric itself (it would be convenient to include examples for greater clarity; thinking that one has done very well looking at the rubric and then the grade shows the opposite; difficult to agree to see which points had been fulfilled and which have not; limitation for requirement for everything to be perfect; it does not have maximum preciseness; always oriented toward the final result and never to the development; limits the work itself)	8	29.6	-	-

Limitation of the script itself (answering yes or no is rather categorical since one may not be sure; it would be convenient to include examples; it does not have maximum preciseness)	-	-	5	21.7
The tool is not clear (sometimes you don't know what to answer; there are very generic parts; not knowing how to do what you are asked to do; uncertainty in case you do not know how to do most things)	3	11.1	4	17.4
It takes more time to do the task	4	14.8	2	8.7
Little use (sometimes one forgets the tool; only looked at at the beginning and at the end)	2	7.4	3	13.0
One more of the many tasks that we students have to perform	1	3.7	3	13.0
Does not correspond to what the teacher asks for in class	0	0	1	4.3
Not used to using this tool	0	0	1	4.3
Use of the tool	43	100	38	100
Easy (steps to follow; just pay attention)	30	69.8	32	84.2
Intermediate (some sentences that are not understood; it is necessary to have a clear understanding of the concepts of the course)	6	14.0	2	5.3
Difficult (interpreting what is requested)	4	9.3	1	2.6
Other negative ideas (lack of habit in their use; forgetting to use them; the tool does not contribute much; the tool was only used at the end to review; the tool takes work and time)	3	7.0	3	7.9
Anxiety/stress	35	100	42	100
Does not affect	20	57.1	25	59.5
Increases anxiety/stress (identify sections that have not been done and that you do not understand what is required or how to do the task; makes evident the lack of knowledge about the course; the	8	22.9	9	21.4

tool causes confusion; the tool makes people think that the work is more difficult than it really is; at the beginning it increases confusion because students do not know how to use the tool)				
Decreases anxiety/stress (by being able to review; by identifying that nothing is missing)	7	20	8	19
Co-create the tool with the teacher	36	100	37	100
It would be useful for learning and doing the task better (identify what is most important; have clearer objectives; clarify doubts; be clear about what is being assessed; it would make it easier to use the tool and it would have been used more from the beginning)	16	44.4	8	21.6
Positive (without specifying reasons in detail)	8	22.2	6	16.2
Negative (for not investing time in the process of co-creation; for not knowing how to design the tool, etc.)	4	11.1	5	13.5
The teacher prepares the tools well (she knows what she wants to ask; useful guidelines for students)	8	22.2	11	29.7
Difficult because of my (the student's) level on the course/contents	0	0	7	18.9

Note. The subcategories with the different statements of the students are included in parentheses. Note that there may be more than one subcategory in a student's response and therefore the total number of responses may be higher than the number of students who answered these open-ended questions. The categories have been ordered from highest to lowest percentage.

Supplementary Material E

Most representative responses of the open-ended questions on students' perceptions of the rubrics/scripts

The main benefit, stated by 82.2% of the participants of the rubric group and 95.7% of the script group, was related to self-regulation and some of its phases: “[the script] gives you the steps to follow in the task and you can always refer back to them if you forget something” (#6. Script group). Another benefit, such as obtaining a higher grade, was only stated by students in the rubric group (11.1%): “it is a guide to getting a better grade” (#2. Rubric group). The main disadvantage was the limitation of the personal way of doing the assignment. Up to one-third of the disadvantages exposed in the rubric group indicate this limitation of the rubric: “limit yourself to what the rubric conveys and do not go deeper, thinking that the rubric is the only thing that helps” (#10. Rubric group). In addition, students identified limitations of the rubric, as being “limited to everything going perfectly” (#41. Rubric group); and in the case of scripts “having to answer categorically yes or no seems to me to be a bit limited as you may not be certain” (#6. Script group). Next, 69.8% of the statements about difficulty in the rubric group, and 72.7% in the script group, stated that it was easy to use the tool. As for how the use of the tool affects task stress, slightly more than half of the statements in both conditions stated that it does not affect it. About 20% stated that it increased: “it is a bit overwhelming to have to self-assess what I am doing, especially if the task is very difficult” (#39. Rubric group); and another 20% stated that it decreased: “because I have a guide to rely on and I can check if I am missing something or if I have not understood it clearly” (#14. Rubric group). With regard to designing the tool together with the teacher, almost half in the rubric group (44.4%) and 21.6% in the script group stated that it would be useful for learning and doing the task better: “could be useful to

understand better or to add things that we [the students] think about” (#25. Rubric group).

Around 20% of the statements even stated that it would have been a positive practice, although the students did not state precisely why they had this opinion. In addition, around 25% in both conditions stated that it is enough that the teacher designs the tools since she is the one who could do it best: “the teacher is the one who knows the objective that we have to achieve with our work” (#26. Rubric group). Finally, in the last question about whether they wished to add something else, the students reported some interesting ideas (among others repeated), mainly regarding performing the task in a group. Although most of their suggestions were expressed on only one occasion, this does not detract from their value. Representative examples are especially significant, such as the following with respect to group collaboration: “my teammates noticed things in the task that I did not, even though I had used the script” (#6. Script group); “I was able to share my doubts with the group” (#23. Rubric group); “by doing it among the whole group you learn new things” (#24 & #26 Script group); “we all agreed on the same thing [when checking the script]” (#29. Script group); “maybe I have not had a good working group” (#37. Rubric group). In addition, students also revealed other situations experienced: “in the first practice we only saw the rubric at the end. Then before, during and after” (#17. Rubric group); “we would have done the work the same way with or without the script” (#14. Script group); “I did not know how to use the rubrics efficiently” (#3. Rubric group).