Assessment in Education: Principles, Policy & Practice

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/caie20

Using rubrics for formative purposes: identifying factors that may affect the success of rubric implementations

Anders Jonsson, Ernesto Panadero, Leire Pinedo & Belén Fernández-Castilla

To cite this article: Anders Jonsson, Ernesto Panadero, Leire Pinedo & Belén Fernández-Castilla (01 Apr 2025): Using rubrics for formative purposes: identifying factors that may affect the success of rubric implementations, Assessment in Education: Principles, Policy & Practice, DOI: 10.1080/0969594X.2025.2486947

To link to this article: https://doi.org/10.1080/0969594X.2025.2486947

Published online: 01 Apr 2025.

ASSESSMENT IN EDUCATION: PRINCIPLES, POLICY & PRACTICE
https://doi.org/10.1080/0969594X.2025.2486947

Using rubrics for formative purposes: identifying factors that may affect the success of rubric implementations

Anders Jonsson [a], Ernesto Panadero [b], Leire Pinedo [c] and Belén Fernández-Castilla [d]

[a]Faculty of Education, Kristianstad University, Kristianstad, Sweden; [b]Centre for Assessment Research Policy and Practice in Education (CARPE), School of Policy and Practice, Institute of Education, Dublin City University, Dublin, Ireland; [c]Facultad de Educación y Deporte, Universidad de Deusto, Bilbao, Spain; [d]Metodología de las Ciencias del Comportamiento y de la Salud, ,Universidad Nacional de Educación a Distancia, Madrid, Spain

ABSTRACT

The formative use of rubrics seems to have the potential of promoting student learning, supposedly by making expectations and criteria explicit. However, there is a variation in effects on how well students perform on academic tasks when supported by rubrics. The aim of this study was to identify factors in rubric interventions that may potentially explain this variation in effects. This was investigated by analysing 15 'high-quality studies' reporting on rubric interventions. The 'success' of these studies was ranked, based on the effect size on academic performance from rubric interventions. We performed a content analysis, searching for similarities and differences in procedures and instrumentation. Our analysis revealed two key characteristics common to successful interventions: First, teachers explain both the content of the rubric and its application for formative purposes. Second, an effective sequence involves students writing or producing work, followed by feedback or self-assessment, and subsequent revision.

ARTICLE HISTORY
Received 22 December 2023
Accepted 26 March 2025

KEYWORDS
Critical review; instructional design; rubrics; student performance; systematic review

Rubrics are widely regarded as valuable tools for supporting student learning. By making expectations and criteria explicit, the use of rubrics may facilitate feedback and self-assessment (Panadero & Jönsson, 2013). However, research findings on such formative use of rubrics are not uniform (Panadero et al., 2023). While some studies report notable improvements in student performance, others show minimal or no benefits. These inconsistencies indicate that the impact of rubrics may depend on various moderating factors, such as the way rubrics are designed, introduced, and integrated into the learning process. Yet, we currently have limited insights into what these factors might be, and how they influence the outcome of rubric interventions (Brookhart, 2018; Panadero et al., 2023). This study seeks to address this gap by identifying and analysing key characteristics of rubric implementations that may explain the observed variation in effects on student academic performance.

CONTACT Anders Jonsson anders.jonsson@hkr.se Kristianstad University, SE-291 88 Kristianstad, Sweden

A definition of rubrics

A rubric is an instrument that 'articulates expectations for student work by listing criteria for the work and performance level descriptions across a continuum of quality' (Brookhart, 2018, p. 1). Rubrics are sometimes also described as combining 'three essential features: evaluative criteria, quality definitions, and a scoring strategy' (Popham, 1997, p. 72). However, rubrics designed for formative purposes may not necessarily include an explicit scoring strategy (see e.g. Wiggins, 1998). In this study, we will therefore not use the term 'scoring rubrics', as this implies the existence of a scoring strategy, which not all rubrics have.

The most prevalent format for representing rubrics involves a table or matrix structure. In this layout, the initial column on the left typically outlines the assessment criteria. Subsequent columns accommodate various performance levels, which can exhibit a range from high to low quality, or conversely, low to high quality. The top row often features labels for these performance levels (such as 'outstanding', 'deficient', etc.), and in the case of 'scoring rubrics', it may specify the points associated with each level. There are also less commonly employed designs that incorporate these same elements, but utilise alternative visual representations, such as concentric circles.

In his seminal work on the use of rubrics for formative purposes, Wiggins (1998) suggests that rubrics provide transparency to the assessment situation by making criteria and performance levels explicit. This information can then be used by the students to facilitate their self-regulation and/or feedback processes, which in turn can have an impact on their academic performance.

Effects of the use of rubrics on students' performance

Rubrics have the potential of promoting student learning, supposedly by making expectations and criteria explicit (Jonsson & Svingby, 2007; Panadero & Jönsson, 2013). By drawing on empirical work about the formative use of rubrics in higher education, Jonsson (2020) identifies four lines of research supporting the idea that rubrics may help students regulate their learning and improve their task performance. First, students may use rubrics to understand expectations. This is seen in studies investigating students' perceptions of being provided with rubrics (e.g. Reynolds-Keefer, 2010), where the results suggest that students both appreciate being provided with rubrics and are able to understand and use the criteria. Secondly, students may use rubrics to plan, monitor, and evaluate their task performance. This is seen in studies investigating how students' use, or claim to use, rubrics (e.g. H. Andrade & Du, 2005).

A third line of research reports on improved task performance due to the use of rubrics. In these studies, students are provided with a rubric either before task performance, so that they can use the rubric to plan and monitor their performance, or afterwards so that they can use the rubric to revise/improve their performance (e.g. Howell, 2014; Kocakülah, 2010). In contrast to studies on students' understanding or use of rubrics for self-regulation, the studies in this category generally rely on quantitative comparisons rather than self-reported data. It is this kind of research that forms the basis of the meta-analysis performed by

Panadero et al., (2023), which the present study builds upon, showing that the use of rubrics generally has a positive influence on students' academic performance.

Finally, there is a line of research reporting on the influence on students' self-efficacy and goal orientations due to the use of rubrics. However, this research is not as straight forward as in the previous categories. For example, findings from research on the impact on student self-efficacy from using rubrics is inconclusive (Panadero & Jönsson, 2013), and too few quantitative studies were found that could be included in the meta-analysis by Panadero et al., (2023). Furthermore, a meta-analysis on the effects of self-assessment interventions on SRL strategies showed that students using rubrics reported *lower* self-efficacy after the intervention than participants not using them (Panadero et al., 2017).

Taken together, there is substantial evidence supporting the claim that the use of rubrics may have a positive impact on student self-regulation and performance. However, as seen in the meta-analysis by Panadero et al., (2023), as well as in previous reviews (e.g. Brookhart & Chen, 2015), there are also studies reporting on very low or absent effects. Typically, these studies report statistically significant effects for only some criteria or some groups of students. For example, Goodrich Andrade (2001), and Goodrich Andrade and Boulay (2003), conducted studies on writing performance, where the students were supported by a rubric and self-assessment. In these studies, positive effects are reported only for some of the groups investigated, and the results sometimes differ between genders.

The overarching question guiding this study is why there is a large variation in effects when using rubrics to support students' academic performance.

A meta-analysis on the formative use of rubrics

In the first systematic review on the use of rubrics (Jonsson & Svingby, 2007), it was noted that rubrics seem to have the potential of promoting student learning. Also, it was suggested that rubrics may support student learning by making expectations and criteria explicit, which in turn facilitates formative practices, such as feedback and self-assessment. However, due to a small number of studies ($n = 25$), no strong claims could be made about the formative potential of rubrics at that time.

More recent reviews have drawn similar conclusions as Jonsson and Svingby (2007) about the formative potential of rubrics. However, it has also noted that there is a scarcity of high-quality studies on the topic, and that available reviews are mainly narrative in their approach. It has therefore not been possible to establish or quantify the effects of using rubrics on student performance (e.g. Brookhart, 2018; Brookhart & Chen, 2015; English et al., 2022; Panadero et al., 2013; Reddy & Andrade, 2010).

The meta-analysis by Panadero, Jönsson, et al. (2023) was performed to, in some sense, settle the debate on the potential merits of the formative use of rubrics. The analysis, encompassing 54 standardised mean differences (k = 21, 95% CI [0.312, 0.831]) and 30 standardised mean changes (k = 12, 95% CI [0.02, 0.75]), showed a clear positive effect of using rubrics on students' academic performance (effect sizes of 0.45 and 0.38 respectively). These findings suggest that the formative use of rubrics may substantially support students' academic performance.

Despite evidence showing that the formative use of rubrics can improve student performance, findings are not uniform. Some studies report robust positive effects,

while others find only modest effects, or no improvement at all. Initially, one might attribute these discrepancies to well-known demographic or contextual factors, such as student gender or age. However, findings from the meta-analysis (Panadero et al., 2023) have failed to establish clear, systematic relationships between these commonly studied moderators and the variability in rubric effects. In other words, even though variables like age and gender have been considered, they have not sufficiently accounted for the observed range of outcomes. This lack of explanatory power suggests that other factors – possibly related to how rubrics are introduced, explained, integrated into instruction, or aligned with assignments – may influence the results. Identifying these factors is central to gaining a deeper understanding of why rubrics sometimes yield substantial benefits, but at other times do not. The present study aims to address this gap by examining high-quality studies with varying effect sizes, to uncover patterns in the implementation and design of rubric interventions that may explain the inconsistencies reported in the literature.

Critique against the use of rubrics for formative purposes

In addition to research suggesting that rubrics may facilitate student learning, there are critics arguing against the use of explicit criteria for formative purposes. For example, Sadler (2009) argues against the use of analytical assessments and pre-set criteria, which are typical characteristics of rubrics, since analytic assessments focus on the parts rather than the whole, which may involve a risk of fragmentation. Instead, he advocates holistic assessments with 'emergent' criteria, which means that assessors should not set any criteria beforehand, but address criteria that surface in the moment of assessing a particular piece of work. The main argument for this approach is the 'indeterminacy' of criteria, which suggests that no matter how many discrete components holistic judgements are broken down into, and no matter how carefully they are selected, these components cannot sufficiently represent the full complexity of the multi-criterion qualitative judgement made by the assessor.

There are also arguments against sharing rubrics with students. For example, Sadler (2014) argues that students may not understand the criteria, since words, symbols, diagrams, and other 'codifications' lack the necessary attributes to represent the criteria or standards. Any attempt to communicate academic expectations through such means is therefore 'fundamentally flawed'. According to Sadler, students therefore need to develop a conceptualisation of what constitutes 'quality' by continuously evaluating authentic work, without being hampered by pre-set criteria.

Torrance (2007, 2012) also claims that the use of explicit criteria may turn students' attention away from productive learning and focus on surface strategies and 'criteria compliance' instead. The core aspiration of education should therefore be on students' autonomous thinking, rather than on the convergent thinking produced by pre-set criteria.

Despite the growing critique against the use of rubrics, there is a general lack of empirical evidence to support several of the claims made by critics. Instead, the claims are mainly supported by anecdotal evidence and/or personal experiences. This means that we currently have a limited and incomplete picture of the potentially negative effects of using rubrics for formative purposes (Panadero & Jönsson, 2020).

Rubric design and implementation

If the main route for rubrics to support student performance and self-regulation is through communicating expectations, rubrics intended to be used formatively need to be both designed and implemented in a way that supports this purpose. In a scoping review on the use of rubrics in the K-12 context, English et al. (2022) present two lines of research in relation to rubric design. First, the use of rubrics alone, without any support materials or explicit training, is insufficient for K-12 students (cf. e.g. Panadero & Jönsson, 2020). Second, teachers' capacity to write rubrics that describe levels of quality can be a limitation of using rubrics, and the design of rubrics can be a possible reason for the lack of effects in some studies. Overall, however, the authors claim that the rubrics used in the studies reviewed were not sufficiently investigated to draw any conclusions about the effect of the design.

A study by Jonsson (2014) discusses features of rubric design and implementation to facilitate students' understanding and use of the criteria more thoroughly. This study is based on an investigation of students' perceptions and uses of rubrics in three different assessment situations, where rubrics were used to convey assessment expectations to students. The features of rubric design and implementation identified are to make the rubrics *accessible* and *aligned* with the assignments (cf. visual, procedural, and linguistic accessibility; Graham et al., 2018).

In this context, the term 'accessible' refers to students' understanding of the rubric and how to use it, as well as availability. This means that for the rubric to be accessible, the teacher needs to explain the meaning of the criteria in the rubric. In the study by Jonsson (2014), this was done in different ways by the different teachers, suggesting that it is not necessarily how this is done that is important, but that it is done. The teachers in this study also provided access to the rubrics *before* the students carried out their assignments, which made it possible for the students to use the rubrics as guides when planning, executing and evaluating their performance. Furthermore, the teachers made the rubrics available to the students by publishing the documents digitally or by handing them out on paper. As a consequence, the students did not have to rely on the teacher's oral description and/or their own interpretations and notes. Instead, the criteria could be reviewed and discussed both individually and among peers. The students could also have the rubrics beside them when they performed their assignments or, in one of the cases, assessed the performance of peers.

A second feature of rubrics that was seen to facilitate students' understanding and use of the criteria (besides accessibility) was alignment. By being aligned with the assignment at hand, the students were invited to use the rubrics as guides to performance, as well as tools for self-assessment and reflection. Jonsson (2014) specifically address how rubrics may be placed along a continuum from highly task specific to generic as an aspect of alignment, since this may affect how useful a rubric is for students. As also suggested by Jonsson and Panadero (2017), rubrics for formative purposes should not be entirely task-specific, but preferably have an appropriate balance of 'task-level specificity', so that they are applicable to several, but similar, tasks assessing the same set of skills.

Potential moderator variables

Given the extensive use of rubrics, a great variation in design and implementation is to be expected, including different moderating factors that may have an impact, positive or negative, on the effects of using rubrics. For this reason, Panadero et al., (2023) explored a total of 11 variables as possible moderators on the effect of using rubrics on academic performance. These variables were (1) year of publication, (2) student gender, (3) mean age, (4) educational level, (5) school or higher education, (6) number of sessions to implement and use the rubric, (7) number of assessment criteria, (8) number of performance levels in the rubric, (9) if the rubric is used in combination with students' self and peer assessment, (10) research design, and (11) quality of the study. However, none of these moderators were significantly related to the observed effect sizes in the meta-analysis on standardised mean differences. As described above, mean age and educational level significantly predicted the variability of the standardised mean changes, but in different directions.

Due to the inconclusive findings, secondary analyses were carried out, using the standardised mean differences to explore the relationship between the number of sessions and mean age; variables that have been proposed by Panadero et al., (2013) to interact. The results from these analyses show that the length of the intervention has less impact on the observed effect in samples with a larger mean age (i.e. longer interventions may be needed for younger students and vice versa).

Needless to say, the lack of statistical significance in these analyses does not rule out a possible relationship between moderators and effects. There are several reasons for this. First, the number of studies exploring the different moderators was low, which means that the analyses lacked statistical power. Second, the level of detail on the intervention characteristics was minimal in several studies. Panadero, Jönsson, et al. (2023) therefore provide an instrument to support the report of rubric interventions and research design characteristics for future studies. Third, in interventions performed in authentic classrooms, the use of rubrics is often coupled to other educational interventions. Taken together, this means that a plausible explanation of the observed variation in effect sizes reported by rubric intervention studies is lacking, making it interesting to explore what makes rubric interventions successful from a qualitative point of view.

Aim and research questions

Our aim is to identify factors in rubric interventions that may potentially explain the variation in effects when using rubrics to support students' academic performance. The research questions (RQ) are:

RQ1. What characteristics are common to more successful implementations of rubrics in the sample of studies used for this study?

RQ2. Which factors may explain less successful implementations of rubrics?

Materials and methods

Selection of studies

The selection of studies in this study is based on a recent meta-analysis exploring the effects of rubrics on academic performance, self-regulated learning, and self-efficacy between the years 2000 to 2022 (Panadero et al., 2023). The rationale for our selection is that we aim to explore studies that present implementations of rubrics that are more or less successful in relation to academic performance, but only include 'high-quality studies' to extract reliable conclusions. As mentioned previously, other recent reviews have noted that there is a scarcity of high-quality studies on the topic, and are mainly narrative in their approach (e.g. Brookhart, 2018; Brookhart & Chen, 2015; English et al., 2022; Panadero & Jönsson, 2013; Reddy & Andrade, 2010).

Since Panadero et al., (2023) not only reviewed the existing literature but also analysed the research quality of the studies included in the meta-analysis, their data can be used as a solid foundation for selecting experimental and quasi-experimental studies of high quality for this study as well.[1]

Of the 23 studies included in the meta-analysis, 21 provided 54 effect sizes for academic performance and were of potential interest for this study. In order to quantify the quality of rubric intervention studies, Panadero et al., (2023) used a scoring rubric (see Appendix). This rubric contains seven criteria focusing on quality (e.g. research design, quality of measurement instruments, quality of intervention, etc.), and three levels (Weak-Medium-Robust). For any criteria judged as 'Weak', no points are awarded, while criteria judged as 'Medium' or 'Robust' are awarded one or two points, respectively. To qualify as a 'high-quality study', and therefore be included in this study, a cut-off score of 7 points was set, which means that the study had to score at least 'Medium' for all criteria, or compensate any number of criteria judged as 'Weak' with a corresponding number of criteria judged as 'Robust'. This procedure resulted in the inclusion of 15 'high-quality studies', with 40 accompanying effect sizes ranging from −0.51 to 1.62 (mean = 0.49), and a 0.31 correlation between quality and effect size. The sample and effect sizes are presented in Table 1.

Measuring success

The meta-analysis by Panadero et al., (2023) included two measures of 'success': standardised mean differences and standardised mean change. The first measure indicates the magnitude of the difference in student performance between a control group and an intervention group at a specific point in time, while the second is based on pre-, and post measures for both the control-, and the intervention group. A positive result indicates that the performance of the intervention group was higher as compared to the control group (standardised mean differences), or that the outcome of the intervention group has improved to a greater extent (from the pre-, to the post-test) as compared to the control group (standardised mean change).

As can be seen in Table 1, the effect sizes range from −0,51 (a large negative difference/change) to 1.62 (a very large positive difference/change). In our study, we have used the median as a cut-off point for 'successful' and 'less successful' implementations. This means that implementations with effect sizes below 0.411

Table 1. Selection of studies based on quality score (cut-off point = 7). Data retrieved from Panadero et al., (2023).

Author (year)	Educational level	Quality score	Effect size	Sampling variance
Successful implementations				
H. L. Andrade et al. (2008)	Elementary school	10	0.660	0.080
			0.855	0.039
			0.985	0.074
H. L. Andrade et al. (2010)	Middle school	9	0.660	0.026
Bay and Kotaman (2011)	Higher education	9	1.038	0.069
			1.406	0.076
Bradford et al. (2016)[a]	Elementary school	13	0.921	0.132
Giacumo and Savenye (2020)	Higher education	7	0.440	0.031
			0.703	0.032
Lipnevich et al. (2022)	High school	12	0.919	0.044
			1.115	0.046
			1.147	0.046
			1.378	0.049
			1.446	0.050
			1.619	0.053
Panadero et al. (2012)	High school	7	0.632	0.052
			0.705	0.052
			0.937	0.055
Mixed results				
Goodrich Andrade (2001)	Middle school	7	−0.035	0.034
			0.245	0.039
			0.996	0.105
Manzanares et al. (2015)	Higher education	9	0.000	0.034
			0.016	0.034
			0.019	0.034
			0.028	0.034
			0.030	0.034
			0.107	0.034
			0.382	0.035
			0.735	0.036
Less successful implementations				
Bohlin (2000)	Elementary school	11	−0.067	0.023
			−0.051	0.021
Brewer (2002)	Middle school	7	−0.048	0.064
Giacumo et al. (2013)	Higher education	10	−0.181	0.038
			−0.165	0.040
			−0.141	0.038
			0.123	0.042
Montanero et al. (2014)	Elementary-middle school	8	−0.250	0.031
			−0.164	0.031
Smit et al. (2017)	Elementary-middle school	13	0.223	0.005
Sundeen (2014)	High school	11	0.381	0.028

[a]This study included a comparison of two groups of students, who had both used rubrics, but at different occasions. The effect size from this comparison has not been used here, since the implementation was similar in both groups.

are considered 'less successful' and the ones above 'more successful'. Interestingly, our cut-off point is basically the same as proposed by Hattie (2012), where values above 0.4 are considered 'desired' overall effects. This resulted in six studies with all their effect sizes below the cut-off point (Bohlin, 2000; Brewer, 2002; Giacumo & Savenye, 2020; Montanero et al., 2014; Smit et al., 2017; Sundeen, 2014), two studies with mixes results (Goodrich Andrade, 2001; Manzanares et al., 2015), and seven with all effect sizes above the median (Andrade et al., 2008, 2010; Bay & Kotaman, 2011; Bradford et al., 2016; Giacumo et al., 2013; Lipnevich et al., 2022; Panadero et al., 2012).

Analysis

The studies included in the sample represent a broad range of different interventions. For example, the educational level spanned from elementary school to higher education. Although different writing tasks (e.g. 'persuasive essays') were the most prevalent, there were also other assignments, such as the construction of multiple-choice tests (for pre-service teachers) and mathematical reasoning. As these different interventions were scattered across the range of effect sizes, and therefore unlikely to explain the variation, the analysis focused on the more fine-grained details of the intervention procedures.

The following questions, based on previous publications on the design and implementation of rubrics for formative purposes (e.g. Jonsson, 2020; Jonsson & Panadero, 2017), guided the analysis:

- When were the students provided with the rubric (i.e. before, during, or after task performance)?
- Did the teacher explain, or in any other way provide support for, how to interpret and understand the rubric and the criteria?
- Did the teacher explain, or provide examples of, how to use the rubric for formative purposes?
- Did the students receive training in how to use the rubric for formative purposes?
- Were the students given the opportunity to revise their performance with the support of the rubric?
- Were the students provided with feedback on their task performance prior to making revisions?
- Did the design of the rubric provide adequate support for (self-)assessing the assignment and making decisions on how to improve task performance?
- How were self-, and peer assessments connected to the use of rubrics?

First, for the present work, the studies included were read, searching for answers to the above-mentioned questions, which were primarily located in the Procedure sections of the original publications. Extracts from the studies were then compared, in order to identify similarities and differences among them. For example, in some studies, there were descriptions of how the students were explicitly taught about how to interpret and use the rubric, while in other studies the rubrics were only made available to the students. Finally, the similarities and differences were checked against the effect sizes in Table 1, to identify any factors that may potentially explain the difference between more or less successful interventions. *Perhaps just a line or two on the similarities and differences found and the characteristics found.*

Results

An overview of the main characteristics of the implementations is presented in Table 2. However, it is important to keep in mind that this is a very coarse categorisation that do not do justice to the variation and complexity of the different studies. For example, even if most implementations offered students some kind of feedback, there was great variation in the details. In some cases, students received feedback from the teacher, in other cases from peers.

Table 2. Characteristics of rubric implementations.

	Characteristics							
Study	Assignment	When provided with rubric	Explained the rubric	Explanation or training for formative use	Feeback on performance	Opportunity to revise	Supportive rubric	Self- and/or peer assessment
Successful implementations								
H. L. Andrade et al. (2008)	Writing (language)	Before	Yes	Yes	Yes	Yes	Yes	Self-assessment
H. L. Andrade et al. (2010)	Writing (language)	Before	Yes	Yes	Yes	Yes	Yes	Self-assessment
Bay and Kotaman (2011)	Construction of multiple-choice tests	Before	Yes	Yes	Yes	Not clearly stated	To some extent	Self- and peer assessment
Bradford et al. (2016)	Writing (language)	Before	Yes	Yes	Yes	Yes	Yes	Self-assessment
Giacumo and Savenye (2019)	Online discussion forum performance	Before	Yes	Yes	Yes	Yes	Yes	Self- and peer assessment
Lipnevich et al. (2022)	Writing (language)	After first draft	Yes	Yes	No	Yes	Yes	Self-assessment
Panadero et al. (2012)	Landscape analysis	Before	No	Yes	Yes	Yes	Yes	Self-assessment
Mixed results								
Goodrich Andrade (2001)	Writing (language)	Before	Yes	Not clearly stated	No	Yes	One not student friendly	No
Manzanares et al. (2015)	Test on acquired knowledge in Structural Engineering.	Before	Not clearly stated	Not clearly stated	Not clearly stated	Not clearly stated	Yes	Self-assessment
Less successful implementations								
Bohlin (2000)	Science open-response tasks.	Before	Yes	Yes	Yes	Yes	To some extent	Self-assessment
Brewer (2002)	Expository writing in science	Before	Yes	Yes	Yes	Yes	To some extent	Self- and peer assessment
Giacumo et al. (2013)	Online discussion forum performance.	Before	No	No	Not clearly stated	Not clearly stated	Yes	No
Montanero et al. (2014)	Writing (language)	Before	Yes	Yes	Yes	Yes	Yes	Peer-assessment[a]
Smit et al. (2017)	Mathematical reasoning	Before	Not clearly stated	Yes	Yes	No	To some extent	Self- and peer assessment
Sundeen (2014)	Writing (language)	Before	No/Yes[a]	Not clearly stated	Yes	Yes	Yes	No

[a]Depending on condition.

In some cases, the use of self-, and peer assessment was an integrated part of the intervention, in other cases self-evaluation support was present, but voluntary to use. No conclusions can therefore be drawn from the table only. Below, the findings from the analysis are presented and supported by extracts from the studies.

RQ1. What Characteristics are Common to Successful Implementations of Rubrics?

We identified two main characteristics, namely 'Explaining the rubric' and 'Rubrics used in an instructional sequence', which we present below.

Explaining the rubric

A prevalent feature of studies reporting on successful interventions is that the teacher (or someone else) clearly explains the meaning of the content in the rubric. For example, Lipnevich et al. (2022), who introduced a writing rubric for high-school students, write that:

> During the first session (T1), teachers introduced assignments and discussed the list of the criteria delineated in rubrics for each of the three categories of reading, writing, and analysis, but without the detailed description of the four performance levels within each rubric category. (p. 139)

In this study, the teachers also:

> / . . . /showed students how to use rubrics and exemplars to self-assess and revise their work. The teachers discussed the usage of rubrics and exemplars and had a hands-on demonstration of how students may go about using these tools. (p. 139)

As is obvious from these excerpts, the students in this study received a very thorough introduction to the rubric and how to use it. Similarly, in a successful implementation by H. L. Andrade et al. (2008), elementary school students were taught how to use the rubric to self-assess their work:

> / . . . /students were asked to underline key phrases in the rubric with colored pencils (e.g. 'clearly states an opinion'), then underline or circle in their drafts the evidence of having met the standard articulated by the phrase (e.g. his or her opinion). If they found they had not met the standard, they were asked to write themselves a reminder to make improvements when they wrote their final drafts. This process was followed for each criterion on the rubric except the conventions criterion, which was not formally self-assessed. (pp. 6-7)

The information on how the rubric was implemented differ substantially between the studies, which means that it is not always possible to know to what extent the students received instructions on how to understand or use the rubric. In one of the less successful implementations, however, the students received a rubric without any explicit instruction on how to use it:

> A rubric was posted in half of the course section forum links for student access as a hard scaffold in the discussion board. (Giacumo et al., 2013, p. 4)

Rubrics used in an instructional sequence

Another prevalent feature of studies reporting on successful interventions is that the rubric is used in a sequence of writing/producing, receiving feedback or self-assessing, and revising. For example, in a successful implementation by Bay and Kotaman (2011) preservice teachers were asked to write multiple-choice test items, and the qualities of a good test item were discussed using a rubric. After learning about the qualities of good test items, the students prepared multiple-choice tests, which were evaluated by themselves, their peers, and the lecturer.

Another example is provided by Panadero et al. (2012). In their successful implementation, secondary school students entered their solutions to assignments as text into a computer, received feedback regarding their performance, and repeated the procedure for two more tasks.

In contrast to the studies mentioned, which clearly outline a pedagogical sequence of production and revisions (i.e. 'guided practice'), a number of studies with mixed results or less successful implementations do not mention any such instructional sequences (Giacumo & Savenye, 2020; Giacumo et al., 2013; Manzanares et al., 2015), report on conditions that resemble a standardised testing environment (Sundeen, 2014), or do not provide a sufficiently detailed description of procedure to draw any conclusions about such instructional sequences (Smit et al., 2017). In one less successful implementation, peer feedback was used instead of teacher feedback (Montanero et al., 2014).

RQ2. Which Factors may explain less Successful Implementations of Rubrics?

Two factors that may explain less successful implementations of rubrics were identified, which we have called 'Inadequate support by the rubrics' and 'Differences in task difficulty'.

Inadequate support by the rubrics

As described above, in relation to the study by Jonsson (2014), alignment with the assignment is an important feature of rubrics in order to facilitate students' understanding and use of the rubric. In the analysis, two forms of 'misalignment' have been identified.

First, there are two studies in the sample that focus on student performance on science tasks, and that represent less successful implementations. In both cases, the rubrics only provide support for the procedural or structural parts of the task performance, but not the subject knowledge, which may vary from task to task. This means that even if the students understand the criteria and are able to use them when writing their answers to the tasks, the criteria do not provide any clear guidance on how to perform well on the science content. The effect on student performance from using the rubric is therefore only partial.

To exemplify, in the study by Bohlin (2000), fourth grade students' performance on open-response items in science were investigated. In one of the tasks, the students were asked to explain if the same amount of sugar will dissolve more quickly in water if the sugar was in the form of cubes or as loose crystals. In another task, the students were

asked to compare different leaves and give examples of similarities and difference. For all tasks, students' performances were assessed with the support of a rubric with the following criteria: science words, comprehensibility ('All parts are there'), number of details, and that the answer should 'make sense'. Although these are reasonable qualities when writing answers to science tasks, students' performance on these tasks also depends on subject knowledge (e.g. knowledge about how to dissolve substances in water, and important characteristics of leaves), which varies with the specific tasks. This means, as mentioned above, that the support provided by the rubric is only partial.

The situation in the study by Brewer (2002), which investigated the use of a rubric for teaching and assessing expository writing in science for junior high students, is very similar. In this study, the rubric contains criteria about restating the question, using appropriate examples, giving detail and support, answering the question completely, and providing the reader with 'full understanding', which may be considered adequate for the procedural or structural parts of the task performance, but they do not provide guidance on scientific content knowledge.

It is possible that the less successful implementation reported by Smit et al. (2017) can be explained by the same partial support provided by the rubric. In this study, the rubric encompasses aspects of reasoning skills in mathematics, which also represent procedural or structural parts of task performance, but no criteria to assess more specific subject knowledge. In this case, however, there is not enough information provided to know how students' performance was assessed in detail.

A second form of 'misalignment' identified in the analysis is the use of rubrics that are not 'student friendly'. For example, in the less successful implementation by Montanero et al. (2014), the authors write that there 'is no doubt that the rubric's format is difficult for a Primary student to interpret'. (p. 197), which could potentially explain the low effect size reported. Similarly, in the study by Goodrich Andrade (2001), there was no observed effect on the first assignment (of three), which could be explained by the fact that the rubric (according to the author) was not written in a particularly student-friendly way, while the second and third rubrics were written in more accessible language.

Differences in task difficulty

There are a number of studies in the sample where the students used different rubrics for slightly different tasks or prompts, resulting in (sometimes quite large) variations in student performance. For example, in the study by Goodrich Andrade (2001), middle school students were asked to write three different essays approximately 1 month apart. Results show that there was a positive effect of providing students with rubrics on the second essay, but not the first or third. According to the author, there are several plausible explanations for this pattern, such as (as already mentioned) the rubric for the first essay not being written in a particularly student-friendly way, but it is also suggested that the different kinds of essays may require different kinds of writing skills. For instance, it may be the case that historical fiction essays are more difficult to write, as compared to autobiographical essays.

The study by Panadero et al. (2012), who investigated secondary school students' analyses of three landscapes, provides a second example. The students in this study analysed a landscape while thinking-aloud, submitted their final written analysis into the

same computer where they were watching the landscape, and received feedback regarding their performance. After the feedback, the students analysed a second landscape and the procedure was repeated, and then again for a third landscape. The students in the rubric condition outperformed the control group at all three occasions, and both groups received a lower score on the third assignment. This pattern is not commented by the authors, but it is stated in the Materials section that the difficulty increased for the assignments, so that the third landscape was the most challenging task. It should be noted, however, that there was a second treatment condition, where the students used a so-called 'script' instead of a rubric, which did not demonstrate the same pattern (the highest performance was on the third assignment).

A third example is provided by Manzanares et al. (2015), who used a total of six different rubrics for evaluating the knowledge of students in the same number of topics in Structural Engineering. As can be seen in Table 1, there is great variation among the effect sizes, which range from 0 to 0.735 for the same intervention.

Summary of findings

The analysis of the studies reviewed suggests that two important characteristics are common among the studies reporting on successful interventions (i.e. provide effect sizes above 0.411). First, the teacher (or sometimes someone else) explains the meaning of the content in the rubric, as well as how to use the rubric for formative purposes. Second, the rubric is used in a sequence of writing/producing, receiving feedback or self-assessing, and revising. Studies reporting on less successful interventions generally lack a proper sequence of production and revision as described in the second point, or they do not provide a sufficiently detailed description of procedure to draw any conclusions about such instructional sequences.

The analysis of the studies reviewed also suggests that some instances of less successful interventions can be explained by rubrics not being properly aligned with the task, either by only providing partial support for the assignment or by not being 'student friendly'. In these cases, the rubrics may fail to facilitate student understanding and use of the rubrics for improved performance and self-regulation. In fact, several effect sizes in this category were negative. In some studies, there was a variation among the reported effect sizes for the same intervention, suggesting a difference in difficulty for the tasks used.

Discussion

The aim of this study was to identify factors in rubric interventions that may potentially explain the variation in effects when using rubrics to support students' academic performance. This was investigated by analysing a sample of 'high-quality studies' reporting on rubric implementations, searching for similarities and differences in procedures and instrumentation.

One clear message in relation to RQ1 is that the use of rubrics is more powerful when implemented in relation to a pedagogical intervention framework. That successful interventions have a thorough implementation, and clearly integrate the use of rubrics in an instructional sequence, resonates well with previous research (e.g. Panadero & Jönsson, 2013). Jonsson (2014) uses the term 'accessible' to turn

our attention to students' understanding of the rubric, where the teacher needs to explain the meaning of the criteria in the rubric, as well as the availability of the rubrics, where the students need to have access to the rubric *before* they carry out their assignments, so that they can use it to plan, monitor, and evaluate their performance.

It has also been pointed out repeatedly (e.g. English et al., 2022; Panadero & Jönsson, 2020) that simply handing out rubrics to students does not necessarily lead to desired outcomes, even if some students (especially in higher education) are obviously able to understand and use the rubrics with minimal guidance and still have learning benefits out of this use (e.g. Panadero & Romero, 2014). Although the length of the intervention has less impact on the observed effect in samples with a larger mean age, as shown in the meta-analysis by Panadero et al., (2023), more thorough implementations may still provide a better ground for understanding and using the rubrics, resulting in larger gains among the students. The findings from this study support this view, by showing that the use of rubrics within an instructional sequence is a common characteristic among high-quality studies reporting on successful rubric interventions.

The findings in relation to RQ2 partly reinforce the abovementioned findings, as some of the studies reporting on less successful implementations fail to provide the accessibility called for, by using rubrics that the students are not able to fully understand.

However, even though the alignment between the rubric and the assignment is emphasised as an important feature of rubric design and implementation (Jonsson, 2014), the idea that the low effect sizes in rubric implementations can be partially explained by rubric effects being confounded with content knowledge, has not been thoroughly addressed in the previous research (cf. English et al., 2022). This may to some extent be an effect of the predominance of interventions focusing on writing skills in language subjects. Although students' knowledge about the topic of writing is likely to have an impact on the outcome, it is still the writing skills that dominate. Furthermore, in most cases, the same set of transferable skills are used in different tasks. This is expressed in terms of 'task-level specificity' by Jonsson and Panadero (2017), when discussing how to design rubrics for formative purposes. Task-level specificity means that the rubric is neither too closely tied to a particular task, which may be used only once, nor too generic (cf. Dawson, 2017). Instead, the rubric is applicable to several, but similar, tasks assessing the same set of skills. This is partly the case with the science rubrics used in the studies by Bohlin (2000) and Brewer (2002), which include criteria on (for example) comprehensibility and detail. However, the subject knowledge required for providing details on solving sugar in water differs radically from the knowledge needed to compare characteristics of leaves, which means that training to use the rubric in one domain of science does not easily transfer to another. A similar argument can be made for the mathematical reasoning skills in the rubric used by Smit et al. (2017). Consequently, if using rubrics with tasks that rely heavily on students' content knowledge, the variation due to different content needs to be taken into consideration. Otherwise, successive tasks are not comparable.

Differences in task difficulty pose a similar problem. If comparing student performance on several tasks, the relative difficulty of these tasks needs to be known, otherwise the difficulty of the task may either mask or exaggerate the effect of using a rubric.

Educational implications

The major educational implication from this review is to clearly plan the implementation of rubrics by explaining the meaning of the content in the rubric, as well as how to use the rubric for formative purposes. Even if the findings from the meta-analysis by Panadero et al., (2023) suggest that the length of the intervention has less impact on the observed effect with older students, and some students in higher education may be able to use the rubric without support, the gains are likely to be higher if the rubric is used within an instructional sequence. As can be seen in Table 1, several effect sizes from higher education are comparatively low.

Furthermore, the rubric should be used in a sequence of writing/producing, receiving feedback or self-assessing, and revising. The details of how to follow these recommendations are likely to depend on the context (e.g. subject, task, etc.), making it difficult to provide any generic guidelines. However, the studies reporting on successful implementations provide several exemplary descriptions that can be used as inspiration and/or scaffolding structures when designing new implementations.

Another important implication is to carefully consider what kind of knowledge and skills that the rubric provides support for, as an important dimension of alignment between rubric and assignment. As suggested by Jonsson and Panadero (2017), rubrics for formative purposes should have an appropriate balance of 'task-level specificity', so that they are applicable to several, but similar, tasks assessing the same set of skills. Less transferable content knowledge may not be equally suitable to include when using rubrics formatively. Also, as noted by English et al. (2022), teachers' capacity to write rubrics that describe levels of quality can be a limitation of using rubrics, which means that schools may need to provide teachers with training on how to write rubrics, so that the criteria are both sufficiently transparent to the students and aligned with the qualities to be assessed.

Limitations and future research

This is a critical and systematic review, based on a qualitative analysis of a selection of studies explored in a meta-analysis. This approach may be considered innovative, as the qualitative analysis is anchored to the quantitative results of the meta-analysis. Due to the qualitative methodology, it has been possible to analyse the texts in detail and identify patterns that were not possible to detect with the meta-analytic approach. On the other hand, the review is based on a limited sample of high-quality experimental and quasi-experimental studies drawn from a recent meta-analysis. As a result, only certain factors emerged as consistently associated with successful rubric implementations. It is important to note that other strategies documented in the broader literature (such as co-constructing rubrics or using rubrics to evaluate sample essays) may also play significant roles in supporting student learning. However, these approaches were not present in this selected sample. Future studies with broader inclusion criteria or different research designs may identify additional factors that further explain the variation in rubric effects.

A particularly important limitation of this study is that the analysis is based on the available text only. As some articles lacked key information about the rubric design and implementation, we cannot be sure whether some aspects were considered and

implemented in the research or not. This calls for caution when interpreting the findings, and we recommend treating them primarily as hypotheses for future research.

Another limitation of this review is the potential presence of publication bias. It is possible that certain high-quality studies were never published due to the absence of significant effects, potentially biasing the sample of studies we intended to analyse and thereby affecting the validity of the results. It should also be noted that the instrument used to assess the quality of the studies has not yet been thoroughly validated, which means that some measurement error could be expected.

As a recommendation for future research, we would specifically like to turn the attention to the design features of rubrics that have been identified as potential contributions to less successful implementations, since they (as compared to the length and quality of interventions) have not been thoroughly addressed in previous research. Consequently, it would be interesting to see future studies disentangle the effects of students' content knowledge, as well as task difficulty, and the skills measured by the rubric.

Conclusions and contributions

The analysis suggests that successful rubric interventions have a thorough implementation of the rubric and clearly integrate the formative use of the rubric in an instructional sequence. That the rubric is not properly aligned with the task, or that the rubric is not adapted to the students who are supposed to use it, are suggested as potential explanations for the particularly low effects reported in some studies. Differences in task difficulty could also explain why students perform lower than expected on some occasions during the same implementation.

These findings provide a refined perspective on the conditions under which rubrics are most effective, contributing to the existing literature in several ways. First, while previous research has acknowledged that rubrics can improve performance, this analysis clarifies why some interventions yield substantial gains while others do not. By pinpointing two core implementation characteristics (i.e. explicit explanations of rubric use and integrating rubrics into an instructional sequence involving feedback and revision), we move beyond general recommendations and offer specific actionable insights. Second, we introduce a nuanced understanding of how certain contextual factors, such as misalignment between rubric criteria and the nature of the assignment, may undermine intended benefits. This adds depth to the current literature, which has often focused on broad moderators like student demographic variables, without fully accounting for the instructional ecology of rubric use. Thus, our study does not merely reiterate known findings; rather, it advances the conversation by explaining some of the variability in previous results and providing guidance for researchers and practitioners aiming to optimise rubric implementation.

Finally, the methodology, where the qualitative analysis is anchored to the quantitative results of the meta-analysis can also be considered an important contribution of the study, making it possible to analyse the data in detail and identify patterns that were not possible to detect with the meta-analytic approach.

Note

1. The data is available at: https://osf.io/d6mry/?view_only=0db0f9c5617a492c84500c0a83a185b9

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Anders Jonsson is Professor of Education and Director of Doctoral Studies at Kristianstad University (Sweden). His research focuses on assessment and grading.

Ernesto Panadero is Professor and Director of CARPE (Centre for Assessment Research, Policy and Practice in Education) at Dublin City University (Ireland) and Ikerbasque Professor and international leader of ERLA (Education, Regulated Learning, & Assessment group) at Deusto University (Spain). His research focuses on self-regulated learning and educational assessment.

Leire Pinedo is a Ph.D. student at ERLA (Education, Regulated Learning, & Assessment group) at Deusto University (Spain). Her research focuses on self-assessment.

Belén Fernández-Castilla is Professor at the National University of Distance Education in Madrid (Spain). Her research is in the field of psychometrics.

ORCID

Anders Jonsson http://orcid.org/0000-0002-3251-6082
Ernesto Panadero http://orcid.org/0000-0003-0859-3616
Leire Pinedo http://orcid.org/0000-0002-3046-5226
Belén Fernández-Castilla http://orcid.org/0000-0002-3451-0637

References

Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation*, *10 (3)* 3. https://doi.org/10.7275/g367-ye94

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

*Andrade, H. L., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement Issues & Practice*, *27*(2), 3–13. https://doi.org/10.1111/j.1745-3992.2008.00118.x

*Bay, E., & Kotaman, H. (2011). Examination of the impact of rubric use on achievement in teacher education. *New Educational Review*, *24*(2), 283–292. https://doi.org/10.15804/tner.11.24.2.22

*Bohlin, S. L. (2000). *Effectiveness of instruction in rubric use in improving fourth-grade students' science open-response outcomes* (Publication No. 304607287) [Doctoral dissertation]. University of Massachusetts Lowell, ProQuest Digital Dissertations database.

*Bradford, K. L., Newland, A. C., Rule, A. C., & Montgomery, S. E. (2016). Rubrics as a tool in writing instruction: Effects on the opinion essays of first and second graders. *Early Childhood Education Journal*, *44*(5), 463–472. https://doi.org/10.1007/s10643-015-0727-0

*Brewer, D. (2002). *Teaching writing in science through the use of a writing rubric (Publication No. 1760593502)* [Doctoral dissertation]. University of Michigan-Flint, ProQuest Digital Dissertations database.

Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education*, *3* (22). https://doi.org/10.3389/feduc.2018.00022

Brookhart, S. M., & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review*, *67*(3), 343–368. https://doi.org/10.1080/00131911.2014.929565

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

English, N., Robertson, P., Gillis, S., & Graham, L. (2022). Rubrics and formative assessment in K-12 education: A scoping review of literature. *International Journal of Educational Research*, *113*, 101964. https://doi.org/10.1016/j.ijer.2022.101964

*Giacumo, L. A., & Savenye, W. (2020). Asynchronous discussion forum design to support cognition: Effects of rubrics and instructor prompts on learner's critical thinking, achievement, and satisfaction. *Educational Technology Research & Development*, *68*(1), 37–66. https://doi.org/10.1007/s11423-019-09664-5

*Giacumo, L. A., Savenye, W., & Smith, N. (2013). Facilitation prompts and rubrics on higher-order thinking skill performance found in undergraduate asynchronous discussion boards. *British Journal of Educational Technology*, *44*(5), 774–794. https://doi.org/10.1111/j.1467-8535.2012.01355.x

*Goodrich Andrade, H. (2001). The effects of instructional rubrics on learning to write. *Educational Theory and Practice Faculty Scholarship*, *6*. https://scholarsarchive.library.albany.edu/etap_fac_scholar/6

Goodrich Andrade, H., & Boulay, B. A. (2003). Role of rubric-referenced self-assessment in learning to write. *Journal of Educational Research*, *97*(1), 21–30. https://doi.org/10.1080/00220670309596625

Graham, L. J., Tancredi, H., Willis, J., & McGraw, K. (2018). Designing out barriers to student access and participation in secondary school assessment. *The Australian Educational Researcher*, *45*(1), 103–124. https://doi.org/10.1007/s13384-018-0266-y

Hattie, J. (2012). *Visible learning for teachers*. Routledge.

Howell, R. J. (2014). Grading rubrics: Hoopla or help? *Innovations in Education and Teaching International*, *51*(4), 400–410. https://doi.org/10.1080/14703297.2013.785252

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

Jonsson, A. (2020). Rubrics as a tool for self-regulated learning. In P. Grainger & K. Weir (Eds.), *Assessment rubrics in higher education* (pp. 25–40). Cambridge Scholars Publishing.

Jonsson, A., & Panadero, E. (2017). The use and design of rubrics to support assessment for learning. In D. Carless, S. Bridges, C. Chan, & R. Glofcheski (Eds.), *Scaling up assessment for learning in higher education* (pp. 99–111). Springer.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, *2*(2), 130–144. https://doi.org/10.1016/j.edurev.2007.05.002

Kocakülah, M. S. (2010). Development and application of a rubric for evaluating students' performance on Newton's laws of motion. *Journal of Science Education and Technology*, *19* (2), 146–164. https://doi.org/10.1007/s10956-009-9188-9

*Lipnevich, A. A., Panadero, E., & Calistro, T. (2022). Unraveling the effects of rubrics and exemplars on student writing performance. *Journal of Experimental Psychology Applied*, *29*(1), 136–148. https://doi.org/10.1037/xap0000434

*Manzanares, M. C. S., Báez, M. Á. S., Ortega-López, V., & Villalaín, J. M. M. (2015). Self-regulation and rubrics assessment in structural engineering subjects. *Education Research International*, *2015*, 1–8. https://doi.org/10.1155/2015/340521

*Montanero, M., Lucero, M., & Fernández, M. J. (2014). Iterative co-evaluation with a rubric of narrative texts in Primary Education/Coevaluación iterativa con rúbrica de textos narrativos en

la Educación Primaria. *Infancia Y Aprendizaje*, *37*(1), 184–220. https://doi.org/10.1080/02103702.2014.881653

Panadero, E., Alqassab, M., Fernández-Ruiz, J., & Ocampo, J. C. (2023). A systematic review on peer assessment: Intrapersonal and interpersonal factors. *Assessment & Evaluation in Higher Education*, *48*(8), 1053–1075. https://doi.org/10.1080/02602938.2023.2164884

Panadero, E., & Jönsson, 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., & Jönsson, A. (2020). A critical review of the arguments against the use of rubrics. *Educational Research Review*, *30*, 100329. https://doi.org/10.1016/j.edurev.2020.100329

Panadero, E., Jönsson, 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., Jönsson, A., Pinedo, L., & Fernández-Castilla, B. (2023). Effects of rubrics on academic performance, self-regulated learning, and self-efficacy: A meta-analytic review. *Educational Psychology Review*, *35*(4). https://doi.org/10.1007/s10648-023-09823-4

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

*Panadero, E., Tapia, J. A., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning & Individual Differences*, *22*(6), 806–813. https://doi.org/10.1016/j.lindif.2012.04.007

Popham, W. J. (1997). What's wrong-and what's right-with rubrics. *Educational Leadership*, *55*(2), 72–75.

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, *35*(4), 435–448.

Reynolds-Keefer, L. (2010). Rubric-referenced assessment in teacher preparation: An opportunity to learn by using. *Practical Assessment, Research & Evaluation*, *15*(1), 8.

Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, *34*(2), 159–179. https://doi.org/10.1080/02602930801956059

Sadler, D. R. (2014). The futility of attempting to codify academic achievement standards. *Higher Education*, *67*(3), 273–288. https://doi.org/10.1007/s10734-013-9649-1

*Smit, R., Bachmann, P., Blum, V., Birri, T., & Hess, K. (2017). Effects of a rubric for mathematical reasoning on teaching and learning in primary school. *Instructional Science*, *45*(5), 603–622. https://doi.org/10.1007/s11251-017-9416-2

*Sundeen, T. H. (2014). Instructional rubrics: Effects of presentation options on writing quality. *Assessing Writing*, *21*, 74–88. https://doi.org/10.1016/j.asw.2014.03.003

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. 1. *Assessment in Education*, *14*(3), 281–294. https://doi.org/10.1080/09695940701591867

Torrance, H. (2012). Formative assessment at the crossroads: Conformative, deformative and transformative assessment. *Oxford Review of Education*, *38*(3), 323–342. https://doi.org/10.1080/03054985.2012.689693

Wiggins, G. (1998). *Educative assessment. Designing assessments to inform and improve student performance*. Jossey-Bass.