



# Increased Explicitness of Assessment Criteria: Effects on Student Motivation and Performance

Andreia Balan<sup>1\*</sup> and Anders Jönsson<sup>2</sup>

<sup>1</sup> Municipality of Helsingborg, Helsingborg, Sweden, <sup>2</sup> Faculty of Education, Kristianstad University, Kristianstad, Sweden

## OPEN ACCESS

### Edited by:

Christopher Charles Deneen,  
RMIT University, Australia

### Reviewed by:

Carmen Tomas,  
University of Nottingham,  
United Kingdom  
Kim Schildkamp,  
University of Twente, Netherlands

### \*Correspondence:

Andreia Balan  
andreia.balan@helsingborg.se

### Specialty section:

This article was submitted to  
Assessment, Testing and Applied  
Measurement,  
a section of the journal  
Frontiers in Education

**Received:** 11 April 2018

**Accepted:** 27 August 2018

**Published:** 25 September 2018

### Citation:

Balan A and Jönsson A (2018)  
Increased Explicitness of Assessment  
Criteria: Effects on Student Motivation  
and Performance. *Front. Educ.* 3:81.  
doi: 10.3389/feduc.2018.00081

The purpose of this study was to investigate the effects of increased explicitness of assessment criteria on students' performance and motivation. Successive levels of explicitness, from feedback based on (implicit) criteria to a combination of exemplars and explicit criteria, were implemented in eight classes at four schools ( $n = 153$  students, 12–13 years old) during four teaching sequences in science. Data was collected on: (a) student performance through knowledge tests, (b) student motivation (self-efficacy, goal orientations, and self-regulation) through questionnaires, and (c) perceived clarity of goals and criteria through “exit tickets.” Findings show that student performance improved from pre-, to post-tests at all schools (effect sizes from 0.82 to 1.38), but not in relation to the level of explicitness. There was also an increase in self-efficacy for low-performing students, but, again, not in relation to explicitness. These changes are instead assumed to be an effect of the formative feedback provided as part of the intervention. The only change related to the level of explicitness, was an increase in self-regulation scores by high-performing students when having access to both exemplars and explicit criteria. Findings therefore suggest that low to medium levels of explicitness in assessment have no discernable effects on students' performance or motivation.

**Keywords:** assessment, criteria, feedback, formative assessment, transparency

## INTRODUCTION

Findings from empirical research, where clear goals and explicit assessment criteria have been shared with students, indicate that increased transparency may positively affect student performance, reduce anxiety, as well as support students' use of self-regulated learning strategies. In particular, the use of rubrics has been seen to decrease the level of “performance/avoidance self-regulation,” which refers to actions motivated by negative emotions, such as anxiety (Panadero and Jönsson, 2013). Furthermore, it is suggested that students' motivation for learning is positively affected by their understanding of learning goals and performance criteria (Ellis and Tod, 2015). Fears voiced against the practice of sharing criteria with students is that students may not understand the criteria or that the use of criteria may turn students' attention away from productive learning toward surface strategies and “criteria compliance” (e.g., Torrance, 2007; Sadler, 2009).

Since there is a lack of studies systematically investigating how students are influenced by the use of explicit criteria, it is currently not fully understood under which circumstances it is productive for student learning and motivation to share explicit assessment criteria. The aim of this study is therefore to investigate the influence of increased explicitness of assessment criteria on student performance and motivation.

## BACKGROUND

According to the widely accepted definition by Sadler (1987) a criterion is:

A distinguishing property or characteristic of anything, by which its quality can be judged or estimated, or by which a decision or classification may be made (p. 194).

Following from this definition, using criteria for assessment purposes is a two tier process. The first stage involves the discernment of these “distinguishing properties” in a text, a presentation, a product, or in any other format used, and the second involves making a judgement about the quality of the performance. This conceptualization of assessment differs markedly from a measurement model of assessment, building on test theory (e.g., Shepard, 2000). For example, by focusing on the quality of products, the assessment is *direct* and does not involve any inferences about students’ latent capabilities in terms of proficiency, knowledge, or competency. Nor does the assessment involve any claims about generalizability of the results. The only claim made is about the merits of the current performance. Another important difference between “assessment-as-judgment” and “assessment-as-measurement” is that in the former case, no scale has to be involved. Criterion-referenced assessment may result in a qualitative judgment about the potential of the particular piece of student work, which may be expressed in terms of strengths and suggestions for development according to the criteria.

The abovementioned characteristics of criterion-referenced assessment are responsible for the potential that such assessments have for students’ learning. First, by focusing on strengths and suggestions for development, criterion-referenced assessments are excellent material for formative feedback. As opposed to test results, which are deeply codified and have to be transformed in order to function as formative feedback, criterion-referenced assessments do not need such a transformation. Second, without a common scale, criterion-referenced assessments are not easily comparable between students, which means that the negative effects of social comparisons associated with grading may be avoided. Third, since the assessment is direct, the base/data for assessment is available to the students, which means that they—with time and practice—should be able to judge the quality of their own or others’ performance.

Yet another possibility provided by criterion-referenced assessments is to communicate the criteria to the students prior to their performance. As suggested by for instance Panadero et al. (2016), students could benefit from being familiar with the criteria during all phases of the self-regulation cycle (e.g., Zimmerman, 2013). They can use criteria to set more realistic goals for the activity during the planning phase, monitor their work during the performance phase, and also self-assess their performance during the evaluation phase. However, in order to communicate the criteria to the students beforehand, the criteria have to be made explicit.

### Explicit and Pre-set Criteria

As pointed out by Sadler (1985), people are constantly engaged in appraisals, without necessarily making reference to any (explicit)

criteria. This observation has two important implications. First, the recognition of quality predates any formulation of explicit criteria, which means that explicit criteria are articulated in retrospect. Second, people cannot be devoid of criteria. When judging the quality of something, be it student performance or something else, people have to rely on some kind of criteria. However, these criteria need not be explicit, but implicit and unspecified. They may also be personal, as opposed to being shared by a particular community of practice. Such “latent criteria” basically exist inside the heads of assessors, who might not even be aware of their conceptions, let alone being able to articulate them. Instead, the criteria emerge in the process of judgment. This model of assessment using latent and emerging criteria is common in appraisals of wine, literature, works of art etcetera, where the criteria are more or less inaccessible to others than the connoisseurs or experts. Criteria are also routinely transmitted from the expert to the novice by joint participation in activities involving evaluative judgment, as opposed to communicating the criteria as (more or less) abstract formulations (Sadler, 1987).

Articulating criteria undoubtedly has its advantages. As explained by Säljö (2005), language gives us the possibility to structure the world around us and focus on what is considered relevant in current practice. Furthermore, once criteria have been formulated linguistically, they can be discussed, critiqued, and (possibly) adapted to new contexts.

However, there are also perils of transforming implicit criteria to explicit. Some problems with explicit criteria have been meritoriously discussed by Sadler (e.g., Sadler, 2009, 2014). For instance, Sadler points out that it does not matter how many criteria you define, they will still not be able to represent the richness and complexity of real world performance. This means that teachers always run the risk of encountering student performance that is judged as high quality, but that does not fit into the predefined set of criteria. As suggested by Klenowski and Adie (2009), this problem may be particularly pronounced for novice teachers, who have been seen to be more prone to use criteria and standards “to the letter,” as compared to experienced teachers who tend to use criteria in a more flexible manner.

That explicit criteria cannot fully represent the richness and complexity of real world performance also means that assessments of different parts or aspects of performance does not necessarily add up to the whole. This is particularly evident in cases where sub-scores from analytical assessments are arithmetically added together into a summary score, possibly resulting in a score not in line with a holistic assessment of overall quality. It should be noted, however, that scoring criterion-referenced assessments (as defined here) is questionable, since it means placing qualitatively different dimensions of performance on the same scale and also making the assessment compensatory (Sadler, 1987). It would be more reasonable to express the outcome of criterion-referenced assessments in terms of strengths and suggestions for development (i.e., a qualitative assessment). In such cases, there does not have to be any conflict between analytic and holistic assessments; rather they may complement each other.

The final peril of transforming implicit criteria into explicit ditto that will be discussed here, is the “fuzziness” of criteria.

Sadler (2009) writes that discrete criteria should be conceptually distinct from one another: “Each criterion is assumed to have an established interpretation that, at least in theory, represents a property that is different from those signified by the other criteria, taken singly or together” (p. 166–167). To make this discussion concrete, we can use a practical and common example: assessing the quality of wine. When assessing the quality of wine, connoisseurs typically refer to the balance, intensity, finish, and complexity of the wine. Without going into details about the meaning of these criteria, it is obvious that they can be used by tasters of wine all over the world in order to make meaningful conversations about the quality of wine. Similar criteria can be found in a number of specialized communities, such as masters assessing the speed, strength, technique, and balance of practitioners in martial arts. In both of these cases, the criteria are “distinguishable properties” that can be discerned by experts in these communities, although not necessarily by outsiders or novices. The word that represent these properties, however, are more or less arbitrary. The property of “balance” in wine could probably be called “even-ness” without losing any of the meaning attached to it, since language does not have the precision to express exactly what we mean. Furthermore, although using the same word, “balance” has a quite different meaning in martial arts. In order to come to know the “true meaning” of a criterion, you must therefore learn how it is used in practice. The important point to be made here, however, is that the arbitrary nature of the words chosen to represent the criteria does not necessarily reflect a similar indetermination of the actual criteria, which consist of a combination of words and accompanying practice.

Taken together, there are both advantages and dangers with articulating latent criteria. By making criteria explicit, they can be communicated and discussed, as opposed to implicit criteria that are hidden in the heads or the practice of experts. If communicated and understood prior to task performance, explicit criteria can be used by students to set goals, as well as to monitor and evaluate their work, which may in turn affect their motivation and task performance. However, in order to understand criteria, they also need to understand the practice to which the criteria belong. The arbitrary words used to represent the criteria will typically not be able to communicate the richness and complexity of the qualities that the criteria refer to. Relying solely on these words, in isolation from practice, therefore run the risk of trivializing the original criteria.

## Explicit Criteria and Student Task Performance

There are different ways to make criteria explicit, but here the focus will be on scoring rubrics. There are two reasons for this choice. First, rubrics are probably the most common way to communicate criteria to students (Dawson, 2017), and, second, rubrics are also used in this study as a means of explicating assessment criteria.

In 2007, Jonsson and Svingby (2007) published a review on the use of scoring rubrics for both summative and formative purposes. Rubrics are instruments for assisting assessors in judging the quality of student performance on open and/or complex tasks, as opposed to drawing conclusions about student proficiency based on the quantity of correct answers. All rubrics

have at least two features in common. First, in order to assist in identifying the qualities to be assessed, the rubric includes information about which aspects or criteria to look for in student performance. Second, in order to assist in judging the quality of student performance, the rubric includes descriptions of student performance at different levels of quality. And by combining these features into a two-dimensional matrix, a rubric has been designed (Jönsson and Panadero, 2017).

What Jonsson and Svingby (2007) found, was that the use of rubrics had the potential of promoting learning and/or improving instruction by making expectations and criteria explicit, which facilitated feedback and self-assessment. However, at that time, the number of studies investigating the formative potential of rubrics was quite limited and the Jonsson and Svingby (2007) review included only 25 studies. Since then, the interest in rubrics has steadily grown. Dawson (2017) writes that the 100th paper mentioning “assessment rubrics” was published in 1997, the 1000th in 2005, and sometime in 2013, the 5000th paper mentioning rubrics was published.

In 2013, a new review on rubrics was published, which focused exclusively on the formative function of rubrics (Panadero and Jönsson, 2013). The findings from this review corroborated the findings from the previous one, by showing that the use of rubrics may provide transparency to the assessment, which in turn may: (a) reduce student anxiety, (b) aid the feedback process, and (c) support student self-regulation; all of which may indirectly facilitate improved student performance. Brookhart and Chen (2014) also note, in a follow-up review on both summative and formative uses of rubrics, that several studies reporting on the effects of rubric use on learning and performance used relatively rigorous designs, such as experiments and quasi-experimental studies.

Since then, a number of empirical studies reporting on positive effects on student performance from the use of rubrics have been published. For example, Lipnevich et al. (2014) used experimental design to compare the effects of standardized feedback: a detailed rubric, exemplars, and a combination of both. Findings show that all three conditions led to significant and strong improvements, with the stand-alone rubric leading to the greatest improvement. Similarly, Greenberg (2015) reports that students using a rubric performed with higher quality as compared to students who did not. It should be noted, however, that several of the studies reporting on improved performance are situated in a higher-education context. As already remarked by Panadero and Jönsson (2013), while studies performed in higher-education contexts tend to report on positive results when providing the students with rubrics, longer and larger interventions are typically needed in order to produce positive results in schools. Time devoted to work with the rubric therefore seems more crucial for younger students and studies only investing a few lessons<sup>1</sup> typically report no, small, or mixed results (e.g., Smit et al., 2017).

Interestingly, although the findings are more unambiguous in the higher-education context, this is also where the most vigorous debate concerning explicit criteria can be found. Typically, critics *a priori* assume that rubric-assisted learning is superficial or

<sup>1</sup>Typically less than five, according to Panadero and Jönsson (2013).

misguided. For instance, Torrance (2012) writes in relation to transparency of expectations:

With respect to the core aspirations of higher education, the issue can be stated very bluntly: Are we trying to get students to jump through pre-specified hoops, by making the nature of those hoops more apparent and encouraging students to better understand how the objectives of a course can be met; or are we trying to get students to think for themselves? (p. 330).

Similarly, Sadler (2009, 2014) argues that the idea to develop explicit descriptions of academic achievement standards is “fundamentally flawed” since words, symbols, diagrams, and other “codifications” lack the necessary attributes to represent the criteria or standards. Any attempt to communicate criteria to students through the use of language (or any symbols) are therefore bound to be futile. Still, as reported by Lipnevich et al. (2014), rubrics:

*.../ forced students to examine what they had done, and look to see how it met the requirements of the task, rather than trying to imitate the exemplar without checking their understanding of the task. .../ the rubric may have called for a more sincere and mindful engagement, which resulted in the student carrying out effective revision practices and thus improving their performance (pp. 551–552).*

Correspondingly, in a study by Jönsson (2014), several students claimed that they used the rubric in order to structure and assess the progress of their work, but it was also shown that some students did not use the criteria when they felt that they did not need to. A plausible explanation for these findings is that codifications provided are sufficient for higher-education students, since they are already familiar with the practice to which the criteria belong, while younger students are not (yet).

Taken together, there is accumulating empirical evidence that explicit criteria may support student performance. This is particularly true for higher-education students, while more comprehensive and long-term interventions are needed for younger students. Furthermore, the empirical support for the claim that the use of explicit criteria leading to superficial learning is weak and the critique is typically based on personal and/or theoretical considerations only (e.g., Kohn, 2006; Wilson, 2006). Contrary to this claim, current research rather supports a notion of students as conscious consumers/users of criteria.

## Explicit Criteria and Student Motivation

Regarding the effects of rubric use on motivation, the most common constructs investigated are self-efficacy and self-regulated learning (SRL). The main rationale for assuming that the use of explicit criteria affects these constructs is that the criteria may support students in gaining a deeper understanding of the requirements of the task at hand, thereby being able to set more realistic goals and more accurately estimate their capacity to perform the task (i.e., improving their self-efficacy). Explicit criteria may also support students in monitoring their task performance and facilitating

reflection about the final product (i.e., self-regulate their learning).

In relation to self-efficacy, the findings from empirical research are mixed, making it difficult to draw any firm conclusions (Panadero and Jönsson, 2013; Brookhart and Chen, 2014). For instance, Andrade et al. (2009) found that self-efficacy increased for a group of students using rubrics, as well as the comparison group, but although the increase was larger in the rubric group, the difference was not statistically significant. Furthermore, there was a significant effect of gender, where the self-efficacy of girls were higher. Another example is the work by Panadero and colleagues, where self-efficacy was affected by the use of rubrics in only in one of three studies (Panadero et al., 2012, 2013; Panadero and Romero, 2014).

In relation to SRL the findings are generally positive, but not necessarily straight forward. As an example, Panadero and his colleagues have performed a number of studies relating to students' learning orientations and SRL. In one of their investigations, they found that the level of SRL strategies was higher in a group of secondary-education students using rubrics, as compared to students in a control group (Panadero et al., 2012). In another study, it was found that scores on a performance- and avoidance-oriented SRL scale decreased for pre-service teachers using rubrics (Panadero et al., 2013). In yet another study, Panadero and Romero (2014) found that a group of pre-service teachers using rubrics scored higher on a learning-oriented SRL questionnaire, as compared to students who were asked to self-assess their work without any instrument to facilitate the self-assessment. Again, performance- and avoidance-oriented SRL scores also decreased significantly in the rubric group.

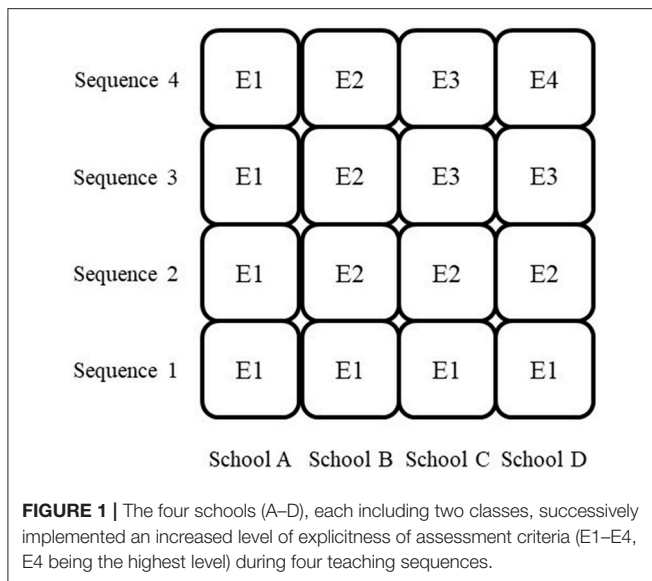
These findings are indeed indications of positive effects on students' SRL, but the students using rubrics in the study by Panadero and Romero also reported higher levels of stress while performing the task as compared to the control group. Furthermore, the learning-oriented SRL scores decreased for psychology students using rubrics (Panadero et al., 2014). This means that while the use of rubrics may decrease performance- and avoidance-oriented SRL strategies, which are often detrimental for learning, they do not necessarily increase learning-oriented SRL.

In sum, research on the consequences of using explicit criteria on students' motivation is still largely under-explored. In particular, given the assumption that access to explicit criteria could foster superficial approaches to learning, it would be imperative to gain a deeper understanding of how students' goal-orientations and other motivational constructs are affected by the use of explicit criteria.

## AIM AND RESEARCH QUESTIONS

As outlined above, the use of explicit criteria has been shown to improve student short-term performance, but mostly in higher-education contexts and maybe also with adverse





consequences for students' long-term learning and motivation. This study therefore aims to investigate the effects of increased explicitness on student performance and motivation in a long-term perspective. Specifically, the study aims to answer the following questions:

1. How is student performance affected by an increase in explicitness?
2. How are students' motivation affected by an increase in explicitness?
3. To what extent are students' perceptions of clarity of goals and assessment criteria affected by an increase in explicitness?

## METHODOLOGY

The overall design of this study is an intervention study, where explicitness of assessment criteria is increased successively over four teaching sequences at four different schools. During the first sequence, all schools taught the same content and used the same level of explicitness. During the second sequence, all schools taught the same content and three schools increased the level of explicitness, while one school remained on the first level. During the third sequence, all schools taught the same content and two schools increased the level of explicitness, while one school remained on the first level and one on the second. During the fourth and last sequence, all schools taught the same content and one school increased the level of explicitness, while one school remained on the third level, one on the second, and one on the first level (Figure 1).

## Sample

The sample in this study is a convenience sample consisting of four primary schools, each including two classes taught by the same teacher. The teachers were found by issuing a call for participation to school leaders in a medium-sized Swedish

**TABLE 1** | Number of students from each school in the sample, as well as school characteristics.

	School A	School B	School C	School D
No. of students in sample	26	49	42	36
No. of students in total	<300	>700	<300	>700
Proportion of immigrant students (%)	>80	<40	<40	<10
Proportion of parents with a higher education degree (%)	<25	>50	>50	>75
Proportion of students passing all subjects (%)	<60	>60	100	100

community, asking for experienced teachers. The participating teachers were selected by their school leaders.

Students in the sample ( $n = 153$ ) attended grade 6 in Sweden, which means that they were 12–13 years old. The number of students at each school can be found in Table 1. Also shown in the table, are some characteristics of the schools, which may influence the results of this study. Note that no exact numbers are presented, since that would make the schools identifiable, as the school statistics are public and available online<sup>2</sup>.

As can be seen in the table, School A is a small school with a high proportion of immigrant students and where the majority of parents lack a higher education degree. Only about half of the students are awarded passing grades in all subjects. Schools B and D, in contrast, are relatively large schools. School D, in particular, differ from School A in having almost no immigrant students, the majority of parents having a higher education degree, and virtually all students leave school with a passing grade in all subjects. School B and C are intermediate in relation to proportion of immigrant students and parents' education. Similar to School D, all students at School C leave school with passing grades in all subjects.

## Procedure

Four teaching sequences were performed during 1 academic year; two during the fall and two during the spring. Each sequence lasted for approximately 3 weeks and before each sequence, the teachers met with the researchers to plan the intervention. First, the researchers described how to implement the different levels of explicitness (see further below). Second, the researchers suggested criteria for assessing students' performance, which were discussed with the teachers and adjusted according to the teachers' suggestions (for an example of the criteria used, see Figure 2). Third, the teachers agreed on the specific content to teach, which they then planned together. This means that for each teaching sequence, the teachers taught the same content, the students performed the same tasks, and the teachers used the same criteria to assess student performance.

During the teaching sequences, students first performed one open-ended task, which was assessed with the criteria and teachers provided formative feedback. The feedback was

<sup>2</sup>School data has been collected from <https://www.skolverket.se/skolutveckling/statistik> (2017-12-14).

Criteria/Level	Level 1	Level 2	Level 3	Level 4
<i>Arguments for and against</i>	Only for or against	Both for and against (one of each)	Several arguments (both for and against)	
<i>Different perspectives</i>	Only one perspective	Only one or several perspectives, but not scientific	Several perspectives, but not scientific	Several perspectives, including a scientific
<i>Justifications</i>	Missing	Own experiences and/or values	Only facts	Makes a distinction between facts and own experiences and/or values
<i>Subject knowledge</i>	Missing	Lacks in relevance and/or correctness	Relevant and correct	

**FIGURE 2** | Example of criteria used for assessment, feedback, and for sharing with students. The criteria in the figure refer to student argumentation in socio-scientific issues.

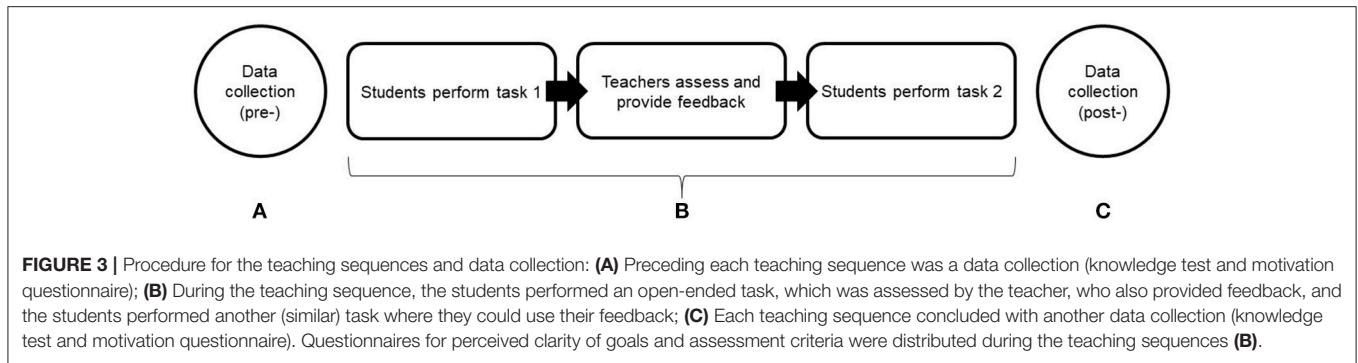
delivered orally to students, either individually, in pairs, or in small groups, depending on how the teachers arranged this purely formative event. Students then performed a similar task (or revised the first one), as an incentive to actively make use of the criteria (**Figure 3**). It is important to note that this process of providing formative feedback and perform a similar assignment (or revise) was identical for *all* sequences and *all* teachers, regardless of condition (E1–E4). It should also be noted that although there were regular meetings and discussions with teachers and researchers, there was no specific training of the teachers.

## Levels of Explicitness

Four levels of explicitness were used in this study and the teachers agreed among themselves which condition they wanted to belong

to. Since all teachers taught two classes, it was initially planned that each teacher should belong to two different conditions—one for each class—in order to compare findings from the same condition with different teachers. This, however, was not considered possible by the teachers for practical reasons. Instead, both classes taught by the same teacher belonged to the same condition.

Although all levels of explicitness implemented (i.e., feedback, exemplars, and explicit criteria) have been shown to generate positive effects on student performance, and hence no students received a negative or neutral intervention, it is important to note that these levels do not necessarily coincide with studies investigating the efficiency of different assessment instruments. For example, in the study by Lipnevich et al. (2014), mentioned above, it was found that the use of a stand-alone rubric led to



greater improvements, as compared to the use of exemplars or a combination of both. Still, explicit criteria are categorized as more explicit than exemplars, and the combination of exemplars and criteria as more explicit than only criteria.

During the first sequence, students were provided with formative feedback based on the criteria (Figure 3), but the criteria were not explicitly shared with the students. The students therefore experienced the criteria indirectly, through the teachers' assessment and feedback. This indirect communication of the criteria was categorized as a low level of explicitness.

During the second sequence, students at three schools were provided with exemplars, chosen to exemplify the criteria (Figure 3). Again, the criteria were not explicitly shared with the students, which means that this was also categorized as a low level of explicitness, but relatively higher as compared to the indirect communication through feedback. According to recommendation from, for example, Panadero et al. (2016), the exemplars were shared with the students prior to performing the task, so that they could use the criteria to inform their planning and goal setting. However, before using the exemplars to support their task performance, the students were given the opportunity to analyze and discuss the exemplars together with the teacher. This discussion focused on identifying the strengths and weaknesses as exemplified by the exemplars, but without making reference to any general or abstract criteria. After this discussion, the students used the exemplars during task performance without teacher assistance. As described above, all teaching sequences involved students solving an open-ended task, which was assessed by the teacher. The formative feedback provided was then used by the students to perform a similar task (or revise the current one). The students therefore actively engaged with the feedback, as well as with the exemplars.

During the third sequence, students at two schools were provided with rubrics, which included explicit criteria. This is therefore the first time during the intervention that the students got the criteria spelled out to them, which was categorized as a high level of explicitness. Students at one school were provided with exemplars, just like during the second sequence, and students at one school only received feedback based on the criteria. Similar to the exemplars, students received the rubrics before they performed the task and they were also given the opportunity to analyze and discuss the criteria with the teacher. They also used the rubric during task performance

without teacher assistance after this discussion. Also similar to the previous condition, the students actively engaged with the feedback, as well as with either the exemplars or rubrics.

Finally, during the fourth sequence, students at one school were provided with both rubrics and exemplars, which is thought to represent the maximum level of explicitness in this study. The remaining students received either a rubric, exemplars, or feedback.

## Teaching Sequences

In the current Swedish national curriculum (Lgr11), the long-term objectives are expressed as "abilities" that the students are supposed to acquire during their time in compulsory school (i.e., grade 1–9). In the natural sciences, there are three such abilities involving (a) communicative aspects of science, (b) systematic investigations, and (c) describing and explaining natural phenomena. Each ability is further concretized by the "knowledge requirements," which are expressed in terms of performance standards (i.e., what the students should be able to do with their knowledge).

In this study, the ability involving communicative aspects of science was chosen for the teaching sequences, since this ability is a relatively new addition to the curriculum and therefore less familiar to the students (i.e., student performance in the study is less affected by previous teaching). In contrast, systematic investigations, as well as describing and explaining natural phenomena, are generally regarded as part of traditional science teaching in Sweden. In the curriculum, there are three aspects of the ability chosen. These are: (i) using knowledge in science in discussions and argumentation, (ii) searching for and reviewing scientific information and different sources, and (iii) using scientific information in text or other representations.

These aspects were used as a framework for teaching by the teachers. For instance, in the first teaching sequence, students were supposed to learn how to use knowledge in science in discussions and argumentation (see criteria in Figure 2). This aspect was combined with specific knowledge in science, in this case sustainable development, where the students learned how to argue about food waste in school. The second teaching sequence focused on information and sources, this time in relation to combustion and pollution. The third teaching sequence focused on using scientific information, where students used written information about forces (like friction) to visualize phenomena

**TABLE 2** | Reliability measures for the knowledge tests.

	Pre-test	Post-test
No. of items	7	7
Alpha	0.712	0.703

with pictures or digital video. The fourth and last sequence again focused on discussions and argumentation, but this time in relation to knowledge about drugs (alcohol, narcotics, and tobacco).

In summary, all teaching sequences were based on the same ability in the national curriculum, but focusing on different aspects and on different content knowledge. The teachers planned the teaching together and used shared plans and assessment criteria for all teaching sequences.

## Data and Data Collection

Data collection was carried out before, during, and after the teaching sequences, which typically had a duration of 3 weeks and were evenly distributed across 1 academic year. Data on student performance was collected with knowledge tests, data on motivation with questionnaires, and data on perceived clarity of goals and assessment criteria with “exit tickets.” Knowledge tests and motivation questionnaires were distributed before and after each teaching sequence, while the exit tickets were distributed during the teaching sequences (**Figure 3**).

The *knowledge tests* were compilations of constructed-response items from previous national tests covering the aspects (i)–(iii) described above. Although tests were distributed after all teaching sequences, only the pre-, and post-tests will be described and reported on here, due to methodological difficulties with the intermediate tests (e.g., low reliability). In order to make the pre-, and post-tests comparable, the tests were calibrated for difficulty by using data (i.e., *f*-values) from the national tests<sup>3</sup>. Furthermore, after initial calibration of criteria, showing satisfactory agreement across raters (Spearman’s rho = 0.943), the tests were scored by a single rater to ensure consistency. Reliability measures (Cronbach’s alpha) for the tests are presented in **Table 2**, including the number of items for each test. It should be noted that the knowledge tests were exclusively used by the researchers to track student progress. No feedback from the tests were provided to the teachers or students, in order to avoid any washback effect on the teaching.

The *motivation questionnaire* was an adaptation of the Students’ Motivation toward Science Learning (SMTSL) (Tuan et al., 2005), which includes scales for self-efficacy, performance goals, achievement goals, and self-regulation that are relevant for this study. The questionnaire used Likert-scale items with six levels (Strongly disagree—Strongly agree). Reliability measures (Cronbach’s alpha) for the scales are presented in **Table 3**,

<sup>3</sup>One item was used as a reference point, while scores from all other items were multiplied by a number depending on the empirically established *f*-values from the national tests. A difficult item therefore generated a higher score, as compared to an easier item.

**TABLE 3** | Reliability measures for the scales in the motivation questionnaire.

	Self-efficacy	Performance goals	Achievement goals	Self-regulation
No. of items	5	4	5	9
Alpha	0.831	0.676	0.741	0.631

**TABLE 4** | Results from the knowledge tests presented as means for each of the schools (in percent).

	School A	School B	School C	School D
Pre-test	58.2 (13.3)	64.3 (19.0)	79.8 (16.6)	70.7 (16.4)
Post-test	81.4 (29.5)	87.6 (17.9)	105.8 (21.0)	87.1 (23.6)

Standard deviations are given in parenthesis. Since the scores have been calibrated for difficulty, the maximum score exceeds 100 percent.

including the number of items for each scale. Sample items are provided in **Data Sheet 1** in the Supplementary Material.

The *exit-tickets* were a single scale questionnaire focusing on perceived clarity of goals and assessment criteria (6 items; Cronbach’s alpha = 0.789). Similar to the motivation questionnaire, the exit tickets used Likert-scale items with six levels (Strongly disagree—Strongly agree). Sample items are provided in **Data Sheet 1** in the Supplementary Material.

## Analysis

Student performance on the knowledge tests were analyzed using descriptive statistics and pre-, and post-tests were compared with *t*-tests within each school. ANCOVA was used to compare post-test results across schools and across levels of explicitness, using results from the pre-test as covariates. Questionnaire data was analyzed with correlational analyses (Pearson’s *r*) and pre-, and post-tests were compared with ANOVA/ANCOVA to identify potential differences within and across groups. Since it was not possible for teachers to implement different conditions in their classes, and all students on each school therefore had the same teacher and took part in identical teaching sequences, data has not been nested in classes. Instead, both classes on each school have been analyzed together.

## FINDINGS

### Student Performance

**Table 4** shows the mean performance of each of the schools for the knowledge tests. As can be seen, results from the pre-test agree fairly well with the school characteristics presented in **Table 1**. During the intervention all schools improved from the pre-test to post-test. In total, the schools improved their scores between 23 and 40 percent, corresponding to a range in effect sizes from 0.82 to 1.38 (Cohen’s *d*) from the pre-, to the post-test.

**Table 5** shows the outcomes of *t*-test analyses between the pre-, and post-tests. As can be seen, the improvement is statistically significant for all schools. However, the findings do



**TABLE 5 |** Comparisons between pre-, and post-tests for each of the schools presented as *t*-test statistics.

	School A	School B	School C	School D
PreT-PostT	3.056**	7.280***	6.684***	5.301***

\*\**p* < 0.01, \*\*\**p* < 0.001.

**TABLE 6 |** Correlations between variables at the start of the study (*n* = 145).

	Self-efficacy	Performance goals	Achievement goals	Self-regulation	Clarity
Self-efficacy	1	0.008	0.146	0.595**	0.452**
Performance goals		1	0.093	0.108	-0.018
Achievement goals			1	0.204*	0.272**
Self-regulation				1	0.486**
Clarity					1

\**p* < 0.05, \*\**p* < 0.01.

not support the assumption that student performance should improve as the level of explicitness increases. This observation is corroborated by the ANCOVA analyses, which show no significant differences between the schools in terms of level of explicitness. ANCOVA analyses also suggest that it is the low-performing students<sup>4</sup> (regardless of school) that increased their performance the most during the intervention, showing a higher estimated mean as compared to other students on the post-test, if using the pre-test as a covariate.

### Students' Motivation

Initial analyses of motivational variables (including perception of clarity of goals and assessment criteria) for the entire sample showed that the correlations between students' perceptions of explicitness and self-efficacy/self-regulation were moderate to strong (Table 6). These correlations did not change considerably over the intervention period (Table 7). However, a stronger correlation could be identified between students' self-efficacy and both achievement and performance goals. A possible interpretation of this is that students who better understood what they could manage in the science course, were also inclined to set both achievement goals and performance goals. This interpretation is supported by the fact that the correlation between achievement goals and performance goals increased during the study. The correlation between self-regulation and achievement goals also increased during the study, but in this case it is more difficult to conclude whether students who set achievement goals are also more self-regulated learners or vice versa.

Table 8 shows the results from the pre-test questionnaire for the self-efficacy and self-regulation scales for the entire sample. Students generally rated their self-efficacy and perception of self-regulation strategies as relatively high on the pre-test questionnaire (4.09 and 4.38 respectively, on a 6 point scale)

<sup>4</sup>Defined as the students in the lower quartile on the pre-test.

**TABLE 7 |** Correlations between variables at the end of the study (*n* = 145).

	Self-efficacy	Performance goals	Achievement goals	Self-regulation	Clarity
Self-efficacy	1	0.173*	0.257**	0.695**	0.375**
Performance goals		1	0.315**	0.152	0.029
Achievement goals			1	0.351**	0.209*
Self-regulation				1	0.444**
Clarity					1

\**p* < 0.05, \*\**p* < 0.01.

**TABLE 8 |** Results from the pre- and post-test questionnaires for self-efficacy and self-regulation scales (*n* = 145).

	Self-efficacy		Self-regulation	
	Pre-test	Post-test	Pre-test	Post-test
Mean	4.09	4.27**	4.38	4.49*
SD	0.600	0.735	0.514	0.649

\**p* < 0.05, \*\**p* < 0.01.

**TABLE 9 |** Results from the pre- and post-test questionnaires for performance-, and achievement goals (*n* = 145).

	Performance goals		Achievement goals	
	Pre-test	Post-test	Pre-test	Post-test
Mean	3.10	3.42**	5.40	5.45
SD	0.961	1.132	0.645	0.743

\*\**p* < 0.01.

across all schools. These ratings could be expected to increase as the level of explicitness increases, but as can be seen in Table 8, the values are more or less unchanged at the end of the intervention.

In relation to achievement-, and performance goals, students' ratings on the achievement goals scale were substantially higher (5.40), as compared to the performance goals scale (3.10) on the pre-test (Table 9). If the use of explicit criteria would make students more performance oriented (i.e., criteria compliant), this relationship could be expected to change. In the current study, however, students' ratings on the achievement goal scale remain unchanged while the performance goals increased only slightly (from 3.10 to 3.42).

Table 10 shows results from the pre-, and post-test questionnaires for the motivational variables for each of the schools in the sample. There were significant differences between the schools on the pre-test questionnaire and in particular the profiles at School A and School D differed in several respects. While students at School A scored relatively low on self-efficacy and self-regulation, and relatively high on both performance-, and achievement goals, students at School D scored low on performance-, and achievement goals, but high on self-efficacy.

**TABLE 10** | Results from the pre- and post-test questionnaires for motivational variables for Schools A–D.

	School A		School B		School C		School D	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
<b>SE</b>								
Mean	3.57	4.00**	4.13	4.17	4.09	4.27	4.38	4.61
SD	0.752	0.559	0.508	0.669	0.554	0.767	0.447	0.798
<b>PG</b>								
Mean	3.51	3.80	3.26	3.19	3.00	3.69***	2.74	3.20
SD	1.240	1.065	0.989	0.881	0.876	1.193	0.658	1.322
<b>AG</b>								
Mean	5.55	5.30	5.58	5.45	5.39	5.51	5.04	5.46**
SD	0.506	0.887	0.400	0.552	0.657	0.806	0.848	0.825
<b>SR</b>								
Mean	3.94	4.17	4.58	4.40*	4.37	4.53	4.39	4.78**
SD	0.685	0.695	0.455	0.568	0.435	0.632	0.379	0.648

SE, Self-efficacy; PG, Performance goals; AG, Achievement goals; SR, Self-regulation. \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

After the pre-test questionnaire most variables either remained unchanged or changed in a negative direction. Some noteworthy changes in relation to individual schools are:

1. School A, which implemented only explicitness Level 1 during the intervention, but has the lowest socio-economic status of the schools in the sample. Students at this school showed significant gains in self-efficacy from pre-, to post-test, and also increased their perception of setting performance goals at the expense of achievement goals.
2. School D, which implemented the highest level of explicitness during the intervention, has the highest socio-economic status of the schools in the sample. Students at this school showed significant gains in both achievement goals and self-regulation.
3. School C, which showed the largest gains in performance goals.

In most cases, however, results on pre-, and post-test questionnaires were similar. For instance, despite the increase, School A still had the lowest score on self-efficacy, as well as on self-regulation, at the end of the intervention. School A also had the highest score on performance goals, despite the fact that School C substantially increased according to this scale. Furthermore, School D has the highest scores on self-efficacy on both pre-, and post-test questionnaires.

**Table 10** shows only the results from the pre-, and post-questionnaires, but there is not much additional information to gain from the intermediate questionnaires. What could be noted is that the increase in self-efficacy at School A, as well as the increase in achievement goals at School D, appear directly after the first teaching sequence and then the scores remain at a higher level. The increase in performance goals at School C, as well as the increase in self-regulation at School D, on the other hand, do not appear until the post-test questionnaire.

Similar to the situation with student performance, therefore, changes in students' perceptions do not appear to be related to

**TABLE 11** | Results from the questionnaires on perceptions of clarity of goals and assessment criteria for Schools A–D.

	School A		School B		School C		School D	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Mean	4.78	4.69	5.09	5.02	4.77	4.54	4.81	5.00
SD	0.909	0.887	0.780	0.693	0.912	1.112	0.814	0.821

the level of explicitness, except for the self-regulation scores at School D, which increased significantly when the students had access to both criteria and examples. Furthermore, and contrary to the situation with student performance, ANCOVA analyses suggest that it is the scores from high-performing students that change the most from pre-, to post-test.

## Perceived Clarity of Goals and Assessment Criteria

**Table 11** shows results from the clarity questionnaires (“exit tickets”) for each of the schools in the sample. There are no statistically significant differences between the groups at the beginning of the intervention and there are no significant changes over time, neither within nor between the schools. Analysis of individual items suggests that students' perceptions about the usefulness of what they are studying in science changed in a negative direction, but that they better understand why they are working with a specific content.

## DISCUSSION

This study aimed to investigate the effects of increased explicitness on student performance, motivation, and perceived clarity of goals and assessment criteria by gradually increasing the level of explicitness during four teaching sequences in primary science. Results suggests that student performance improved during the intervention, but not in relation to the level of explicitness, and that motivational measures, as well as measures of perceived clarity of goals and assessment criteria, did generally not change during the intervention. These findings are discussed below.

## Effects on Student Performance

From previous research on the relationship between transparency and student performance (e.g., Panadero and Jönsson, 2013; Lipnevich et al., 2014) it could be assumed that an increase in explicitness should result in improved performance. In this study, however, this is only partly the case. Although all schools improved their performance from pre-, to post-test, there is no obvious connection between this improvement and the levels of explicitness. Instead, the overall improvement during the intervention was largest at School A and School B, which had the greatest number of low-performing students of the schools in the sample.

Based on the evaluation of the project with the teachers, the improvement could be assumed to be an effect of novelty, where students encountered content (i.e., argumentation in science)

that differed from previous science teaching, in combination with more effective teaching (i.e., the use of formative feedback). To provide formative feedback that was actually used to improve performance, was—according to the teachers—highly motivating for the students and probably the single most ground-breaking aspect of the intervention for them. Since the positive effects of formative feedback are well known (e.g., Hattie and Timperley, 2007; Shute, 2008), it could of course be called into question why this was not already an established part of the teaching. In any case, the effects of implementing formative feedback may have overshadowed any effects of explicitness in the current study.

Taken together, the increase in explicitness does not seem to have had an impact on student performance, beyond the effect of formative feedback.

### Effects on Students' Motivation

One of the main ambitions of increasing transparency is to support student self-regulation, including their self-efficacy. However, previous research on transparency in relation to student motivation has been mixed regarding self-efficacy (Panadero and Jönsson, 2013; Brookhart and Chen, 2014), as well as regarding the use of criteria to support self-regulated learning (Panadero et al., 2017). This study is no exception, since although self-efficacy increased for all schools, only the changes at School A were statistically significant. The most plausible explanation for this increase is that the low-performing students at this school, who also reported relatively low self-efficacy, experienced higher self-efficacy due to the formative feedback; an effect that is consistent throughout the study. The students at the other schools were generally more high-performing, and reported higher self-efficacy already at the start of the intervention and did not change significantly during the study. Rather, the students at School A were more aligned to these students, in terms of both performance and self-efficacy, at the end of the intervention.

The findings are also inconclusive for the self-regulation variable, which increased for School D, but decreased for School B. Since the change in self-regulation appeared when the students at School D had access to both exemplars and criteria, one possible explanation could be that this combination (and thus high level of explicitness) was needed in order to support student self-regulation, while lower levels of explicitness did not. As was shown in a recent meta-analytic review (Panadero et al., 2017), training in student self-assessment is a strong predictor of improved self-regulation, and without explicit training in self-assessment the level of explicitness may need to be very high to make a difference.

If students were to become more criteria compliant during the course of the intervention, it could be assumed that performance-oriented goals should increase in relation to explicitness. This is not the general case in this study, however, since only one of the schools increased the score on the performance-goals scale during the intervention. Furthermore, this increase is seemingly unrelated to the level of explicitness implemented. It should be kept in mind, however, that the score for achievement goals was very high already at the outset and remained high all the way through the study (i.e., above 5 on a 6 point scale) for all schools in the sample, while the score for performance goals was

substantially lower and only increased significantly for one of the schools in the sample.

Taken together, the only support for any effect from the increase in explicitness is the increase in self-regulation at School D. The increase in self-efficacy is more likely to be an effect of formative feedback, and is also primarily confined to low-performing students, and there is no general increase in performance goals. Furthermore, achievement-goals scores remain very high throughout the intervention.

### Effects on Perceived Clarity

Ideally, students' perception of clarity of goals and assessment criteria would increase during the intervention. According to the questionnaire on perceived clarity of goals and assessment criteria, however, this is not the case for most of the students in the sample. Instead, the scores remain unchanged throughout the study, which suggests that students' perceptions of the clarity of goals and assessment criteria were unaffected by the changes implemented.

## CONCLUSIONS

If communicated and understood prior to task performance, explicit criteria can—at least in theory—be used by students to set goals, as well as to monitor and evaluate their work, which may in turn affect their motivation and task performance. However, in order to understand criteria, students also need to understand the practice to which the criteria belong, while the words used to represent the criteria will typically not be able to communicate the richness and complexity of the qualities that the criteria refer to. In the current study, therefore, the criteria were not only communicated as abstract words, but integrated in teaching sequences where the students were encouraged to actively use the criteria as part of formative feedback (see section “Levels of Explicitness” above).

As suggested by the findings, however, the increase in explicitness did not in itself contribute to improved student performance. Although the findings seem to support previous research on the efficiency of formative feedback (e.g., Hattie and Timperley, 2007; Shute, 2008), as evidenced by the large improvement in student performance from pre- to post-test for all schools in the sample, as well as an increase in self-efficacy for low-performing students, this finding cannot be seen as conclusive due to the lack of a control group. Still, the fact that students at a school with low socio-economic status improved both their performance and self-efficacy with such magnitude, likely due to changes in the feedback practice, is worth considering for follow-up studies.

There is also some tentative evidence for the combination of exemplars and criteria contributing to increased self-regulation for high-performing students, but the main conclusion from this study is that the students in the sample are generally unaffected by the increase in explicitness. This could, on the one hand, be interpreted pessimistically, since the findings do not support the idea of transparency (as implemented in this study) being a panacea for improved performance and motivation. On the other hand, it could also be interpreted optimistically, since increased

explicitness does not seem to give any adverse consequences for student motivation—at least not in relation to the measures investigated here.

## Limitations and Suggestions for Future Research

There are several important limitations to this study that need to be considered when interpreting the findings. First and foremost, although care has been taken to provide as similar conditions as possible for all students, for instance by using the same criteria and tasks and by having the teachers plan their teaching together, the teaching sequences are still likely to differ in several respects. It would therefore be desirable to have the same teacher implement different levels of explicitness in their classes, a design which unfortunately was not possible to implement due to practical reasons, or engaging more teachers.

Second, the students in the sample were quite young, which means that their autonomy and capacity to self-regulate were likely limited as compared to older students, possibly resulting in a more uniform outcome for the motivational variables. A sample of older students may therefore provide a clearer and discernable distribution in relation to the questionnaires, but older students may also, on the other hand, have a stronger performance orientation due to the presence of high-stakes grading and national tests, which may mask any effects of increased explicitness.

Third, the use of formative feedback as an incentive for the students to experience and use the criteria may have contributed to the effect on student performance, which means that it has not been possible to identify any potentially fine-grained effects from increased explicitness. Since it is not advisable to refrain from providing students with formative feedback, future research would need to ascertain that the students are accustomed to basic formative-assessment practices, so that the provision of feedback does not become as revolutionary to them.

Fourth, the only documented indication of the effect of explicitness in this study was on student self-regulation, when the students had access to both exemplars and explicit criteria; a high level of explicitness that was only implemented at one of the schools. This school was also the school with the highest socio-economic status in the sample and the students reported high scores for both self-efficacy and self-regulation, as well as the lowest scores for performance goals, on the pre-test

questionnaire. To investigate the generality of this finding, a high level of explicitness would need to be implemented across a more heterogeneous sample, rather than gradually increasing the level of explicitness.

Taken together, in order to further investigate the impact of explicitness on students' performance and motivation, future research should: (a) engage more teachers, so that more than one teacher is assigned to each level of explicitness; (b) include older students with greater capacity to self-regulate their learning; (c) ascertain that the students are accustomed to basic formative-assessment practices, so that the effects of formative feedback do not overshadow any effects of explicitness; and (d) implement higher levels of explicitness across a broader sample of students.

## ETHICS STATEMENT

This study was carried out in accordance with the ethical guidelines for the Humanities and Social Sciences set out by the Swedish Research Council. The study has not been subjected to review by an ethical committee since, according to Swedish legislation regarding research on human subjects (2003:460), research needs approval from an ethical committee only in cases where personal and sensitive information is handled, when physical interventions are made, or when the subjects may be harmed. In line with this, approval from an ethical committee is not required by the university where the research was conducted. All subjects, as well as their legal guardians, have been informed about the purpose of the research, that their participation is voluntary, and that they can interrupt their participation at any time. Written informed consent have been given by all subjects, as well as their legal guardians, in accordance with the Declaration of Helsinki.

## AUTHOR CONTRIBUTIONS

AJ was the principal investigator, who led the design of the study and performed the literature review. Data collection, analyses, interpretation, and writing the manuscript was done in collaboration between AJ and AB.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2018.00081/full#supplementary-material>

## REFERENCES

- Andrade, H. L., Wang, X., Du, Y., and Akawi, R. L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *J. Educ. Res.* 102, 287–302. doi: 10.3200/JOER.102.4.287-302
- Brookhart, S. M., and Chen, F. (2014). The quality and effectiveness of descriptive rubrics. *Educ. Rev.* 67, 343–368. doi: 10.1080/00131911.2014.929565
- Dawson, P. (2017). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assess. Eval. High. Educ.* 42, 347–360. doi: 10.1080/02602938.2015.1111294
- Ellis, S., and Tod, J. (2015). *Promoting Behavior for Learning in the Classroom*. London; New York, NY: Routledge.
- Greenberg, K. P. (2015). Rubric use in formative assessment: a detailed behavioral rubric helps students improve their scientific writing skills. *Teach. Psychol.* 42, 211–217. doi: 10.1177/0098628315587618
- Hattie, J., and Timperley, H. (2007). The power of feedback. *Rev. Educ. Res.* 77, 81–112. doi: 10.3102/003465430298487
- Jönsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assess. Eval. High. Educ.* 39, 840–852. doi: 10.1080/02602938.2013.875117



- Jönsson, A., and Panadero, E. (2017). "The use and design of rubrics to support AfL," in *Scaling up Assessment for Learning in Higher Education*, eds D. Carless, S. Bridges, C. Chan and R. Glofcheski (Dordrecht: Springer), 99–111.
- Jonsson, A., and Svingby, G. (2007). The use of scoring rubrics: reliability, validity and educational consequences. *Educ. Res. Rev.* 2, 130–144. doi: 10.1016/j.edurev.2007.05.002
- Klenowski, V., and Adie, L. (2009). Moderation as judgement practice: reconciling system level accountability and local level practice. *Curr. Perspect.* 29, 10–28.
- Kohn, A. (2006). The trouble with rubrics. *English J.* 95, 12–15.
- Lipnevich, A. A., McCallen, L. N., Miles, K. P., and Smith, J. K. (2014). Mind the gap! Students' use of exemplars and detailed rubrics as formative assessment. *Instruct. Sci.* 42, 539–559. doi: 10.1007/s11251-013-9299-9
- Panadero, E., Alonso-Tapia, J., and Huertas, J. A. (2012). Rubrics and self-assessment scripts: effects on self-regulation, learning and self-efficacy in secondary education. *Learn. Individ. Diff.* 22, 806–813. doi: 10.1016/j.lindif.2012.04.007
- Panadero, E., Alonso-Tapia, J., and Huertas, J. A. (2014). Rubrics vs. self-assessment scripts: effects on first year university students' self-regulation and performance. *Infancia y Aprendizaje. J. Study Educ. Dev.* 37, 149–183. doi: 10.1080/02103702.2014.881655
- Panadero, E., Alonso-Tapia, J., and Reche, E. (2013). Rubrics vs. self-assessment scripts: effect on self-regulation, performance and self-efficacy in pre-service teachers. *Stud. Educ. Eval.* 39, 125–132. doi: 10.1016/j.stueduc.2013.04.001
- Panadero, E., and Jönsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: a review. *Educ. Res. Rev.* 9, 129–144. doi: 10.1016/j.edurev.2013.01.002
- Panadero, E., Jönsson, A., and Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: four meta-analyses. *Educ. Res. Rev.* 22, 74–98. doi: 10.1016/j.edurev.2017.08.004
- Panadero, E., Jönsson, A., and Strijbos, J.-W. (2016). "Scaffolding self-regulated learning through self-assessment and peer assessment: guidelines for classroom implementation," in *Assessment for Learning: Meeting the Challenge of Implementation*, eds D. Laveault and L. Allal (Dordrecht: Springer), 311–326.
- Panadero, E., and Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education. Princ. Policy Pract.* 21, 133–148. doi: 10.1080/0969594X.2013.877872
- Sadler, R. D. (1985). The origins and functions of evaluative criteria. *Educ. Theory* 35, 285–297.
- Sadler, R. D. (1987). Specifying and promulgating achievement standards. *Oxf. Rev. Educ.* 13, 191–209.
- Sadler, R. D. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assess. Eval. High. Educ.* 34, 159–179. doi: 10.1080/02602930801956059
- Sadler, R. D. (2014). The futility of attempting to codify academic achievement standards. *High. Educ.* 67, 273–288. doi: 10.1007/s10734-013-9649-1
- Säljö, R. (2005). *Lärande och Kulturella Redskap: Om Lärprocesser och det Kollektiva Minnet [Learning and Cultural Tools: About Learning Processes and the Collective Memory]*. Stockholm: Norstedts akademiska förlag.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educ. Res.* 29, 4–14. doi: 10.1177/0022057409189001-207
- Shute, V. J. (2008). Focus on formative feedback. *Rev. Educ. Res.* 78, 153–189. doi: 10.3102/0034654307313795
- Smit, R., Bachmann, P., Blum, V., Birri, T., and Hess, K. (2017). Effects of a rubric for mathematical reasoning on teaching and learning in primary school. *Instruct. Sci.* 45, 603–622. doi: 10.1007/s11251-017-9416-2
- 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. *Assess. Educ. Princ. Policy Pract.* 14, 281–294. doi: 10.1080/09695940701591867
- Torrance, H. (2012). Formative assessment at the crossroads: conformative, deformative and transformative assessment. *Oxf. Rev. Educ.* 38, 323–342. doi: 10.1080/03054985.2012.689693
- Tuan, H.-L., Chin, C.-C., and Shieh, S.-H. (2005). The development of a questionnaire to measure students' motivation towards science learning. *Int. J. Sci. Educ.* 27, 639–654. doi: 10.1080/0950069042000323737
- Wilson, M. (2006). *Rethinking Rubrics in Writing Assessment*. Portsmouth: Heinemann.
- Zimmerman, B. J. (2013). From cognitive modeling to self-regulation: a social cognitive career path. *Educ. Psychol.* 48, 135–147. doi: 10.1080/00461520.2013.794676

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Balan and Jönsson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.