

# TRANSPARENCY IN ASSESSMENT – EXPLORING THE INFLUENCE OF EXPLICIT ASSESSMENT CRITERIA

EDITED BY: Anders Jönsson and Frans Prins  
PUBLISHED IN: Frontiers in Education





# frontiers

## Frontiers Copyright Statement

© Copyright 2007-2019 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-799-1

DOI 10.3389/978-2-88945-799-1

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)



# TRANSPARENCY IN ASSESSMENT – EXPLORING THE INFLUENCE OF EXPLICIT ASSESSMENT CRITERIA

Topic Editors:

**Anders Jönsson**, Kristianstad University, Sweden

**Frans Prins**, Utrecht University, Netherlands



Image: Bychikhin Olexandr/Shutterstock.com

In many schools and higher education institutions it has become common practice to share assessment criteria with students. Sometimes it is required for accountability purposes, at other times criteria are used as a means to communicate expectations to students. However, the idea that explicit assessment criteria should be shared with students has been contested. On the one hand, research has shown that explicit criteria may positively affect student performance, reduce their anxiety, as well as support students' use of self-regulated learning strategies. On the other hand, there

are fears that explicit criteria may have a restraining influence on students' learning, as well as limiting their autonomy and creativity. There are also indications of students becoming more performance oriented, as opposed to learning oriented, when being provided with explicit assessment criteria. Taken together, it is not fully understood under which circumstances it is productive for student learning to share explicit assessment criteria, and under which circumstances it is not. In particular, empirical research on the proposed negative effects of sharing criteria with learners is limited and most fears voiced in the literature are based on individual experiences and anecdotal evidence. In this book, we therefore bring different perspectives on transparency in assessment together, in order to further our understanding of how students are influenced by the use of explicit assessment criteria. A deeper understanding of the influence of explicit assessment criteria on students' understanding of criteria, motivation, and learning is equally imperative for future research and educational practice, both of which need to go beyond individual opinions and convictions.

**Citation:** Jönsson, A., Prins, F., eds. (2019). *Transparency in Assessment – Exploring the Influence of Explicit Assessment Criteria*. Lausanne: Frontiers Media.  
doi: 10.3389/978-2-88945-799-1

# Table of Contents

- 05 Editorial: Transparency in Assessment—Exploring the Influence of Explicit Assessment Criteria**  
Anders Jönsson and Frans Prins
- 08 Appropriate Criteria: Key to Effective Rubrics**  
Susan M. Brookhart
- 20 Transparency or Stimulating Meaningfulness and Self-Regulation? A Case Study About a Programmatic Approach to Transparency of Assessment Criteria**  
Liesbeth K. J. Baartman and Frans J. Prins
- 35 Transparency Isn't Spoon-Feeding: How a Transformative Approach to the Use of Explicit Assessment Criteria Can Support Student Self-Regulation**  
Kieran Balloo, Carol Evans, Annie Hughes, Xiaotong Zhu and Naomi Winstone
- 46 Corrigendum: Transparency Isn't Spoon-Feeding: How a Transformative Approach to the Use of Explicit Assessment Criteria Can Support Student Self-Regulation**  
Kieran Balloo, Carol Evans, Annie Hughes, Xiaotong Zhu and Naomi Winstone
- 48 From "Seeing Through" to "Seeing With": Assessment Criteria and the Myths of Transparency**  
Margaret Bearman and Rola Ajjawi
- 56 Increased Explicitness of Assessment Criteria: Effects on Student Motivation and Performance**  
Andreia Balan and Anders Jönsson
- 69 Learning to See New Things: Using Criteria to Support Pre-service Teachers' Discernment in the Context of Teachers' Relational Work**  
Pernilla Holmstedt, Anders Jönsson and Jonas Aspelin
- 83 Are Assessment Exemplars Perceived to Support Self-Regulated Learning in Teacher Education?**  
Peter R. Grainger, Deborah Heck and Michael D. Carey
- 92 Applying Criteria to Examples or Learning by Comparison: Effects on Students' Evaluative Judgment and Performance in Writing**  
Renske Bouwer, Marije Lesterhuis, Pieterjan Bonne and Sven De Maeyer



# Editorial: Transparency in Assessment—Exploring the Influence of Explicit Assessment Criteria

Anders Jönsson<sup>1\*</sup> and Frans Prins<sup>2</sup>

<sup>1</sup> Kristianstad University, Kristianstad, Sweden, <sup>2</sup> Department of Education, Faculty of Social and Behavioural Sciences, Utrecht University, Utrecht, Netherlands

**Keywords:** self-regulation, transparency, assessment, criteria, rubrics

## Editorial on the Research Topic

### Transparency in Assessment—Exploring the Influence of Explicit Assessment Criteria

In many schools and higher education institutions it has become common practice to share assessment criteria with students. Sometimes it is required for accountability purposes, at other times criteria are used as a means to communicate expectations to students. Although it is generally and widely accepted that explicit assessment criteria should be shared with students, challenges to that assumption have been made. On the one hand, research has shown that explicit criteria may positively affect student performance, reduce their anxiety, as well as support students' use of self-regulated learning strategies. On the other hand, there are fears that explicit criteria may have a restraining influence on students' learning, as well as limiting their autonomy and creativity. Taken together, the question guiding this Research Topic is when, and under which conditions, transparency in assessment is productive for learning. The contributions to this Research Topic vary from conceptual approaches to more empirical oriented intervention studies.

## OPEN ACCESS

### Edited and reviewed by:

Gavin T. L. Brown,  
The University of Auckland,  
New Zealand

### \*Correspondence:

Anders Jönsson  
anders.jonsson@hkr.se

### Specialty section:

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

**Received:** 11 December 2018

**Accepted:** 27 December 2018

**Published:** 15 January 2019

### Citation:

Jönsson A and Prins F (2019)  
Editorial: Transparency in  
Assessment—Exploring the Influence  
of Explicit Assessment Criteria.  
Front. Educ. 3:119.  
doi: 10.3389/feduc.2018.00119

## WHEN IS ASSESSMENT TRANSPARENCY BENEFICIAL?

Brookhart, who performed a review of rubrics in higher education, might claim that whether transparency in assessment is productive for learning would depend on the criteria. If the criteria are true indicators of quality, then they have the potential to support student learning. On the other hand, most rubrics in her study proved to be beneficial for student learning, regardless of design. However, what is considered “beneficial” could be discussed. Even if studies on rubrics may report on improved performance, is it not always clear what kind of knowledge has been assessed. Was it, for instance, memory or conceptual knowledge? Convergent or divergent thinking? Short-term or long-term learning? Also, assessment transparency may yield an increase of student's self-efficacy and self-regulatory skills.

The connection between assessment transparency and student self-regulation is explored by several of the authors in this Research Topic. For example, Baartman and Prins, who performed a case study on meaning making of assessment criteria and standards, argue that detailed criteria may be detrimental for students' self-regulation, because it prevents them from choosing their own learning goals. According to them, transparency should therefore be viewed at the curriculum level, addressing what is expected of students at the end of the curriculum and in working life, and linked to the development of self-regulatory skills.

Balloo et al. on the other hand, argue in their conceptual study against the idea that transparency should necessarily foster “criteria compliance” (Torrance, 2007) and learner instrumentalism; instead they suggest that transparency is essential to promoting students' self-regulatory capacity.

That leaves us with the intriguing question of how we can ascertain that transparency will foster one thing and not the other.

## PROVIDING TRANSPARENCY OF ASSESSMENT CRITERIA

In this Research Topic, a few ways of providing transparency in assessment are described. Examples are to explicitly describe criteria and expectations, to provide exemplars, or to have dialogues about assessment criteria with students.

According to Bearman and Ajjawi, transparency cannot be achieved merely through provision of explicit criteria (and maybe not at all), for instance since explicit criteria cannot capture tacit knowledge. This is to some extent corroborated by Balan and Jonsson, who did not find any clear effect of the level of explicitness of expectations on primary school students' motivation and performance in science. However, in the study by Holmstedt et al. pre-service teachers were able to analyze authentic situations with greater precision and at greater depth with the aid of explicit criteria. Apparently, divergent results are found, although it is not totally clear why the results differ.

One possible explanation for differential effects could be student ability, since low-performing and high-performing students responded quite differently to the intervention in the study by Balan and Jonsson. While the low-performing students increased their self-efficacy and performance quite dramatically, the impact on high-performing students was less pronounced. As suggested by previous research (Jonsson, 2014), high-performing students may even choose to ignore the criteria, since they want to manage on their own.

## CONNECTION TO PRACTICE

Another important distinction that emerges from the studies in this Research Topic is the grounding of criteria in the context of practice. In contrast to a “representational view of criteria” (Ajjawi and Bearman, 2018), where each criterion has one single meaning that does not change in relation to the context or the person who interprets them, a sociocultural view holds that explicit criteria are only “the tip of the iceberg.” The greater part of the criteria is tacit, residing in practice (O'Donovan et al., 2004). Consequently, if detached from the practice to which they belong, criteria run the risk of being trivialized. In the study by Holmstedt et al., for example, students were not only provided with explicit criteria, but also guided in the practice of using them in context.

The connection to practice is also explored by Grainger et al. as well as by Bouwer et al. by using examples of authentic performance. Grainger et al. show that students accessed the exemplars regularly and found them useful in providing detailed guidance; a guidance that went beyond the descriptions of assessment tasks found in course outlines and assessment rubrics. Furthermore, students valued various types of exemplars, a range of quality, and the inclusion of annotated and unannotated versions of exemplars.

Bouwer et al. investigated whether students were better prepared for writing after working with a rubric or through learning by comparison. Although they found no effect of condition on the quality of the written essays, students in the comparative judgment condition provided relatively more feedback on higher order aspects, such as the content and structure of the text, as compared to students in the criteria condition.

## ASSESSMENT TRANSPARENCY IN HIGHER EDUCATION

This brings us back to Brookhart, who showed that most uses of rubrics proved to be beneficial for student learning. However, similar to all empirical contributions in this Research Topic, with the exception of Balan and Jonsson, Brookhart's review only included studies from higher education. As already pointed out by Panadero and Jönsson (2013), there seems to be a difference between higher education and school settings. Whereas most interventions in higher education provide positive outcomes, even with no previous training, effects in schools are typically small, partial, or inconclusive—unless the intervention has a very long duration (i.e., several weeks). Not surprisingly, students in higher education are more skilled at using rubrics for self-regulation (i.e., planning, monitoring, and evaluating their performance). One explanation for this could be that students in higher education are older and more mature, or that they have more training in applying self-regulation strategies. Another explanation could be that higher education has a stronger connection to practice. This is evident for professional education, but may apply equally well for the arts and sciences. When studying geology at university, students not only learn the facts and theories of this subject, but also how to practice geology through both laboratory-, and field work. Criteria are therefore more likely to be considered in their context of practice.

## CONCLUSION

We believe that the contributions of this Research Topic will bring the debate about assessment transparency a step further. The conceptual studies disclose considerations and mechanisms, whereas the empirical studies provide some evidence about how specific interventions have effects in practice. This may ultimately lead to process models concerning the effects of transparency of assessment criteria. For now, the studies point to some important prerequisites for transparency in assessment to be productive for student learning:

- First, criteria should be *indicators of quality*. Such criteria do not tell students what to do or how to do it, but how to discern important dimensions of quality in their own (or others') performance.
- Second, criteria need to be considered *in their context of use*, rather than being considered as generally applicable with a fixed meaning. Exemplars, modeling, and feedback can be used to “anchor” the criteria to a community of practice.

- Third, students may need support to *plan, monitor, and evaluate* their performance with the aid of assessment criteria in order to foster learner autonomy and empowerment. Self-regulation does not necessarily occur spontaneously among students, but is a skill to be taught and learned.

## REFERENCES

- Ajjawi, R., and Bearman, M. (2018). "Problematising "standards": representation or performance?" in *Developing Evaluative Judgement in Higher Education: Assessment for Knowing and Producing Quality Work*, eds D. Boud, R. Ajjawi, P. Dawson, and J. Tai (Oxon; New York, NY: Routledge), 41–50.
- Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assess. Eval. Higher Educ.* 39, 840–852. doi: 10.1080/02602938.2013.875117
- O'Donovan, B., Price, M., and Rust, C. (2004). Know what I mean? Enhancing student understanding of assessment standards and criteria. *Teach. Higher Educ.* 9, 325–335. doi: 10.1080/1356251042000216642
- 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
- 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. Principl. Policy Pract.* 14, 281–294. doi: 10.1080/09695940701591867

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

**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 © 2019 Jönsson and Prins. 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.





# Appropriate Criteria: Key to Effective Rubrics

Susan M. Brookhart\*

Department of Educational Foundations and Leadership, Duquesne University, Pittsburgh, PA, United States

## OPEN ACCESS

### Edited by:

Anders Jönsson,  
Kristianstad University College,  
Sweden

### Reviewed by:

Eva Marie Ingeborg Hartell,  
Royal Institute of Technology, Sweden  
Robbert Smit,  
University of Teacher Education St.  
Gallen, Switzerland

### \*Correspondence:

Susan M. Brookhart  
susanbrookhart@bresnan.net

### Specialty section:

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

Received: 01 February 2018

Accepted: 27 March 2018

Published: 10 April 2018

### Citation:

Brookhart SM (2018) Appropriate  
Criteria: Key to Effective Rubrics.  
Front. Educ. 3:22.  
doi: 10.3389/feduc.2018.00022

True rubrics feature criteria appropriate to an assessment's purpose, and they describe these criteria across a continuum of performance levels. The presence of both criteria and performance level descriptions distinguishes rubrics from other kinds of evaluation tools (e.g., checklists, rating scales). This paper reviewed studies of rubrics in higher education from 2005 to 2017. The types of rubrics studied in higher education to date have been mostly analytic (considering each criterion separately), descriptive rubrics, typically with four or five performance levels. Other types of rubrics have also been studied, and some studies called their assessment tool a "rubric" when in fact it was a rating scale. Further, for a few (7 out of 51) rubrics, performance level descriptions used rating-scale language or counted occurrences of elements instead of describing quality. Rubrics using this kind of language may be expected to be more useful for grading than for learning. Finally, no relationship was found between type or quality of rubric and study results. All studies described positive outcomes for rubric use.

**Keywords:** criteria, rubrics, performance level descriptions, higher education, assessment expectations

A rubric articulates expectations for student work by listing criteria for the work and performance level descriptions across a continuum of quality (Andrade, 2000; Arter and Chappuis, 2006). Thus, a rubric has two parts: *criteria* that express what to look for in the work and *performance level descriptions* that describe what instantiations of those criteria look like in work at varying quality levels, from low to high.

Other assessment tools, like rating scales and checklists, are sometimes confused with rubrics. Rubrics, checklists, and rating scales all have criteria; the scale is what distinguishes them. Checklists ask for dichotomous decisions (typically has/doesn't have or yes/no) for each criterion. Rating scales ask for decisions across a scale that does not describe the performance. Common rating scales include numerical scales (e.g., 1–5), evaluative scales (e.g., Excellent-Good-Fair-Poor), and frequency scales (e.g., Always, Usually-Sometimes-Never). Frequency scales are sometimes useful for ratings of behavior, but none of the rating scales offer students a description of the quality of their performance they can easily use to envision their next steps in learning. The purpose of this paper is to investigate the types of rubrics that have been studied in higher education.

Rubrics have been analyzed in several different ways. One important characteristic of rubrics is whether they are general or task-specific (Arter and McTighe, 2001; Arter and Chappuis, 2006; Brookhart, 2013). General rubrics apply to a family of similar tasks (e.g., persuasive writing prompts, mathematics problem solving). For example, a general rubric for an essay on characterization might include a performance level description that reads, "Used relevant textual evidence to support conclusions about a character." Task-specific rubrics specify the specific facts, concepts, and/or procedures that students' responses to a task should contain. For example, a task-specific rubric for the characterization essay might specify which pieces of textual evidence the student should have located and what conclusions the student should have drawn from this

evidence. The generality of the rubric is perhaps the most important characteristic, because general rubrics can be shared with students and used for learning as well as for grading.

The prevailing hypothesis about how rubrics help students is that they make explicit both the expectations for student work and, more generally, describe what learning looks like (Andrade, 2000; Arter and McTighe, 2001; Arter and Chappuis, 2006; Bell et al., 2013; Brookhart, 2013; Nordrum et al., 2013; Panadero and Jonsson, 2013). In this way, rubrics play a role in the formative learning cycle (Where am I going? Where am I now? Where to next? Hattie and Timperley, 2007) and support student agency and self-regulation (Andrade, 2010). Some research has borne out this idea, showing that rubrics do make expectations explicit for students (Jonsson, 2014; Prins et al., 2016) and that students do use rubrics for this purpose (Andrade and Du, 2005; Garcia-Ros, 2011). General rubrics should be written with descriptive language, as opposed to evaluative language (e.g., excellent, poor) because descriptive language helps students envision where they are in their learning and where they should go next.

Another important way to characterize rubrics is whether they are analytic or holistic. Analytic rubrics consider criteria one at a time, which means they are better for feedback to students (Arter and McTighe, 2001; Arter and Chappuis, 2006; Brookhart, 2013; Brookhart and Nitko, 2019). Holistic criteria consider all the criteria simultaneously, requiring only one decision on one scale. This means they are better for grading, for times when students will not need to use feedback, because making only one decision is quicker and less cognitively demanding than making several.

Rubrics have been characterized by the number of criteria and number of levels they use. The number of criteria should be linked to the intended learning outcome(s) to be assessed, and the number of levels should be related to the types of decisions that need to be made and to the number of reliable distinctions in student work that are possible and helpful.

Dawson (2017) recently summarized a set of 14 rubric design elements that characterize both the rubrics themselves and their use in context. His intent was to provide more precision to discussions about rubrics and to future research in the area. His 14 areas included: specificity, secrecy, exemplars, scoring strategy, evaluative criteria, quality levels, quality definitions, judgment complexity, users and uses, creators, quality processes, accompanying feedback information, presentation, and explanation. In Dawson's terms, this study focused on specificity, evaluative criteria, quality levels, quality definitions, quality processes, and presentation (how the information is displayed).

Four recent literature reviews on the topic of rubrics (Jonsson and Svingby, 2007; Reddy and Andrade, 2010; Panadero and Jonsson, 2013; Brookhart and Chen, 2015) summarize research on rubrics. Brookhart and Chen (2015) updated Jonsson and Svingby's (2007) comprehensive literature review. Panadero and Jonsson (2013) specifically addressed the use of rubrics in formative assessment and the fact that formative assessment begins with students understanding expectations. They posited that rubrics help improve student learning through several

mechanisms (p. 138): increasing transparency, reducing anxiety, aiding the feedback process, improving student self-efficacy, or supporting student Self-regulation.

Reddy and Andrade (2010) addressed the use of rubrics in post-secondary education specifically. They noted that rubrics have the potential to identify needs in courses and programs, and have been found to support learning (although not in all studies). They found that the validity and reliability of rubrics can be established, but this is not always done in higher education applications of rubrics. Finally, they found that some higher education faculty may resist the use of rubrics, which may be linked to a limited understanding of the purposes of rubrics. Students generally perceive that rubrics serve purposes of learning and achievement, while some faculty members think of rubrics primarily as grading schemes (p. 439). In fact, rubrics are not as easy to use for grading as some traditional rating or point schemes; the reason to use rubrics is that they can support learning and align learning with grading.

Some criticisms and challenges for rubrics have been noted. Nordrum et al. (2013) summarized words of caution from several scholars about the potential for the criteria used in rubrics to be subjective or vague, or to narrow students' understandings of learning (see also Torrance, 2007). In a backhanded way, these criticisms support the thesis of this review, namely, that appropriate criteria are the key to the effectiveness of a rubric. Such criticisms are reasonable and get their traction from the fact that many ineffective or poor-quality rubrics exist, that do have vague or narrow criteria. A particularly dramatic example of this happens when the criteria in a rubric are about following the directions for an assignment rather than describing learning (e.g., "has three sources" rather than "uses a variety of relevant, credible sources"). Rubrics of this kind misdirect student efforts and mis-measure learning.

Sadler (2014) argued that codification of qualities of good work into criteria cannot mean the same thing in all contexts and cannot be specific enough to guide student thinking. He suggests instantiation instead of codification, describing a process of induction where the qualities of good work are inferred from a body of work samples. In fact, this method is already used in classrooms when teachers seek to clarify criteria for rubrics (Arter and Chappuis, 2006) or when teachers co-create rubrics with students (Andrade and Heritage, 2017).

## PURPOSE OF THE STUDY

A number of scholars have published studies of the reliability, validity, and/or effectiveness of rubrics in higher education and provided the rubrics themselves for inspection. This allows for the investigation of several research questions, including:

- (1) What are the types and quality of the rubrics studied in higher education?
- (2) Are there any relationships between the type and quality of these rubrics and reported reliability, validity, and/or effects on learning and motivation?

Question 1 was of interest because, after doing the previous review (Brookhart and Chen, 2015), I became aware that not

all of the assessment tools in studies that claimed to be about rubrics were characterized by both criteria and performance level descriptions, as for true rubrics (Andrade, 2000). The purpose of Research Question 1 was simply to describe the distribution of assessment tool types in a systematic manner.

Question 2 was of interest from a learning perspective. Various types of assessment tools can be used reliably (Brookhart and Nitko, 2019) and be valid for specific purposes. An additional claim, however, is made about true rubrics. Because the performance level descriptions describe performance across a continuum of work quality, rubrics are intended to be useful for students' learning (Andrade, 2000; Brookhart, 2013). The criteria and performance level descriptions, together, can help students conceptualize their learning goal, focus on important aspects of learning and performance, and envision where they are in their learning and what they should try to improve (Falchikov and Boud, 1989). Thus I hypothesized that there would not be a relationship between type of rubric and conventional reliability and validity evidence. However, I did expect a relationship between type of rubric and the effects of rubrics on learning and motivation, expecting true descriptive rubrics to support student learning better than the other types of tools.

## METHOD

This study is a literature review. Study selection began with the data base of studies selected for Brookhart and Chen (2015),

a previous review of literature on rubrics from 2005 to 2013. Thirty-six studies from that review were done in the context of higher education. I conducted an electronic search for articles published from 2013 to 2017 in the ERIC database. This yielded 10 additional studies, for a total of 46 studies. The 46 studies have the following characteristics: (a) conducted in higher education, (b) studied the rubrics (i.e., did not just use the rubrics to study something else, or give a description of "how-to-do-rubrics"), and (c) included the rubrics in the article.

There are two reasons for limiting the studies to the higher education context. One, most published studies of rubrics have been conducted in higher education. I do not think this means fewer rubrics are being used in the K-12 context; I observe a lot of rubric use in K-12. Higher education users, however, are more likely to do a formal review of some kind and publish their results. Thus the number of available studies was large enough to support a review. Two, given that more published information on rubrics exists in higher education than K-12, limiting the review to higher education holds constant one possible source of complexity in understanding rubric use, because all of the students are adult learners. Rubrics used with K-12 students must be written at an appropriate developmental or educational level. The reason for limiting the studies to ones that included a copy of the rubrics in the article was that the analysis for this review required classifying the type and characteristics of the rubrics themselves.

Information about the 46 studies was entered into a spreadsheet. Information noted about the studies included country, level (undergraduate or graduate), type (rubric, rating

**TABLE 1 |** Types of rubrics used in studies of rubrics in higher education.

| Type                  | How criteria are considered | Performance level descriptions used descriptive language   | Performance level descriptions included rating-scale language and/or relied on counting occurrences | Total |
|-----------------------|-----------------------------|--|---|-------|
| General Rubrics       | Analytic                    | 1 level – 1<br>3 levels – 3<br>4 levels – 14<br>5 levels – 8<br>6 levels – 2<br>8 levels – 1<br>Total – 29 | 4 levels – 4<br>5 levels – 1<br>7 levels – 1<br>Total – 6   | 35    |
|                       | Holistic                    | 4 levels – 3<br>5 levels – 1<br>Total – 4  | 5 levels – 1  | 5     |
| Task-Specific Rubrics | Analytic                    | 2 levels – 1   |   | 1     |
|                       | Holistic                    | 1 level – 2  |   | 2     |
| Rating Scale          | Analytic                    |  | 5   | 5     |
| Point Scheme          | Holistic                    |  | 3   | 3     |
| Total                 |                             | 36   | 15  | 51    |

*Number of rubrics does not equal number of studies because some studies had more than one rubric.*

*General rubrics are general enough to apply to a family of similar tasks and can be shared with students. Task-specific rubrics apply to just one task and cannot be shared with students. Analytic rubrics consider each criterion separately. Holistic rubrics consider all criteria simultaneously.*

*Rating scales require ratings on criteria using a judgmental scale. Examples include numeric scales (e.g., 1–5), frequency scales (e.g., always-usually-sometimes-never), and evaluative scales (e.g., excellent-good-fair-poor).*

*Point schemes are schemes to score tasks by assigning points to various aspects of students' responses.*

**TABLE 2 |** Reliability evidence for rubrics.

| Study                                    | Level                   | Rubric topic & description  | Sample  | Reliability evidence   |
|--|-------------------------|---|---|--|
| Avanzino, 2010                           | Undergraduate           | Oral communication<br>Analytic rubric with 3 criteria, 3 levels with mostly descriptive plds  | 230 speeches (112 individual, 118 group)  | $\kappa = 0.92$  |
| Britton et al., 2017                     | Undergraduate           | Team-Q Rubric for individual teamwork skills<br>Final version: 5 criteria, each with behavioral descriptions, rated with a 5-level frequency scale (never to always)  | 70 students in a theater history and literature course, 24 of whom gave full consent          | External rater ICC 0.76<br>Research assistants ICC 0.77<br>Peers (4–5 per group) ICC 0.79<br>For revised rubric: internal consistency of self-ratings $\alpha = 0.91$ ; internal consistency of peer-ratings $\alpha = 0.97$ |
| Chasteen et al., 2012                    | Undergraduate           | Physics, electromagnetism<br>Detailed task-specific point schemes for each task   | 103 students in 3 courses (final version), 432 students in 14 courses during test development | $\kappa = 0.41$<br>consistency between criteria $\alpha = 0.82$  |
| Cho et al., 2006                         | Undergraduate, graduate | SWoRD Writing Rubrics<br>Analytic rubric with 3 criteria and 7 levels; Plds were somewhat descriptive but relied on counting (e.g., “all but one argument...”) or rating-scale language                           | 708 students in 16 courses over 3 years from 4 universities                                   | Untrained raters<br>Single rater ICCs 0.17–0.56<br>Multiple rater ICCs 0.45–0.88<br>Compared reliability from student and instructor perspectives  |
| Ciorba and Smith, 2009                   | Undergraduate           | Music – Instrumental and vocal performance<br>Analytic rubric with 3 criteria and descriptive plds at 5 levels  | 28 panels of judges, 359 music students' performances   | inter-judge consistency, median $\alpha = 0.89$  |
| DeWever et al., 2011                     | Undergraduate           | Group work<br>Analytic rubric with 4 criteria and descriptive plds at 4 levels  | 659 students in 2 years, in groups of 8–9 (81 groups)   | Untrained raters<br>Single rater ICCs .33 –0.50 (individual criteria), .50 –0.60 (total score)   |
| Garcia-Ros, 2011                         | Undergraduate           | Oral presentation<br>14 criteria organized into 4 areas. 4 levels (0-3) with descriptive plds   | 64 educational psychology students  | exact agreement = 66%<br>adjacent agreement = 98%<br>$\kappa = 0.36$ exact agreement<br>$\kappa = 0.80$ adjacent agreement<br>median $r = 0.89$  |
| Kocakülal, 2010                          | Undergraduate           | Newton's Laws of Motion problem solving<br>Rubric style point scheme; Analytic rubric with 6 criteria and descriptive plds at 5 levels, but points vary depending on the criterion                                | 153 physics students in 4 classes   | Untrained raters<br>single rater ICCs, 0.14, 0.38<br>multiple rater ICCs, 0.93, 0.98<br>instructor's consistency between 2 forms, median $\alpha = 0.76$   |
| Lewis et al., 2008                       | Undergraduate           | Acute care treatment planning<br>Analytic rubric with 4 criteria and descriptive plds at 4 levels   | 22 students, 5 clinical educators, 1 academic faculty   | Expert raters<br>Single rater ICC = 0.32   |
| Menéndez-Varela and Gregori-Giralt, 2016 | Undergraduate           | Service learning projects<br>2 analytic rubrics. Content: 4 criteria, 4 levels each, w/ descriptive plds. Oral presentation: 5 criteria, 4 levels, descriptive plds except for time                               | 84 history of art students  | Project content $\alpha$ increased from 0.67 (at stage 2 of study) to 0.93 (at stage 3 of study; $\alpha$ for oral presentation skills was 0.77  |
| Newman et al., 2009                      | Graduate faculty        | Peer assessment of faculty teaching<br>Rating scale, 1–5 (excellent through does not demonstrate criterion), on 11 criteria   | 14 resource faculty   | Expert raters<br>Single rater ICC = 0.27 (total score)   |
| Nicholson et al., 2009                   | Undergraduate           | Nurse clinical performance in operating suite<br>Analytic rubric with 12 criteria and descriptive plds at 4 levels. Descriptions required inferences (e.g., “would require some prompting and assistance” p. 75). | 40 pre-op nurses rating 3 videos  | Expert raters<br>Single rater ICCs .51 –0.61<br>Multiple rater ICC = 0.98  |
| Pagano et al., 2008                      | Undergraduate           | Writing (College composition)<br>Analytic rubric, 6 levels with descriptive plds at 3 of the levels (1–2, 3–4, 5–6)   | 6 institutions year 1, 5 institutions year 2  | Adjacent agreement = 74%   |
| Reddy, 2011                              | Graduate                | Business Cases, Business Projects<br>Business case study rubric (4 dim); business project rubric (7 dim), each with descriptive plds at 4 levels  | 35 instructors, 95 business students, 2 institutions  | Exact agreement 0.61–0.99<br>Single rater ICCs 0.90–0.95<br>Multiple rater ICCs 0.71–0.99  |

(Continued)



TABLE 2 | Continued

| Study                       | Level         | Rubric topic & description  | Sample  | Reliability evidence  |
|-----------------------------|---------------|---|---|---|
| Rochford and Borchert, 2011 | Graduate      | Business case analysis<br>Analytic rubric, 10 criteria, organized into 4 "subobjectives" using a 1-5 scale with descriptive plds for 1, 3, and 5. | Case analysis assignments in MBA program capstone course                                | Multiple rater ICC = 0.96   |
| Schamber and Mahoney, 2006  | Undergraduate | Critical thinking<br>5 criteria (for each section of the paper) based on Facione and Facione (1996), with descriptive plds at 5 levels            | 2002, 30 papers; 2003, 30 papers  | Median $r = 0.90$   |
| Schreiber et al., 2012      | Undergraduate | Public Speaking Competence Rubric<br>Analytic rubric with 9 criteria (+2 optional), with descriptive plds at 5 levels                             | Study 1, 5 coders, 45 speeches; Study 2, 3 undergraduate + 1 faculty coder, 50 speeches | Expert raters<br>Multiple rater ICCs 0.91, 0.93   |
| Stellmack et al., 2009      | Undergraduate | Writing APA-style introductions<br>Analytic rubric with 8 criteria with descriptive plds at 4 levels  | 40 papers, 3 researcher/graders   | Interrater agreement<br>exact = 0.37, adjacent = 0.90<br>Intrater agreement<br>exact = 0.78, adjacent = 0.98<br>$\kappa = 0.33$ |
| Timmerman et al., 2011      | Undergraduate | Science writing<br>Analytic rubric with 15 criteria and descriptive plds at 4 levels  | 142 lab reports, 9 trained and 8 'natural' graduate student raters                      | Generalizability for relative decisions = 0.85  |
| Wald et al., 2012           | Graduate      | Reflective writing<br>Analytic rubric with 5 criteria (+1 optional) and descriptive plds at 4 levels  | 10–60 narratives over 5 trials  | Single-rater ICCs 0.51–0.75<br>Inter-judge consistency, median $\alpha = 0.77$  |
| Wallace et al., 2011        | Undergraduate | Astronomy – Cosmology<br>Task-specific, holistic rubrics for each test item, with 5 levels  | 65 responses from 21 students, 9 items  | Exact agreement, overall score = 83%<br>$\kappa = 0.76$ , weighted $\kappa = 0.82$  |

plds, Performance Level Descriptions.

scale, or point scheme), how the rubric considered criteria (analytic or holistic), whether the performance level descriptors were truly descriptive or used rating scale and/or numerical language in the levels, type of construct assessed by the rubrics (cognitive or behavioral), whether the rubrics were used with students or just by instructors for grading, sample, study method (e.g., case study, quasi-experimental), and findings. Descriptive and summary information about these classifications and study descriptions was used to address the research questions.

As an example of what is meant by descriptive language in a rubric, consider this excerpt from Prins et al. (2016). This is the performance level description for Level 3 of the criterion Manuscript Structure from a rubric for research theses (p. 133):

All elements are logically connected and keypoints within sections are organized. Research questions, hypotheses, research design, results, inferences and evaluations are related and form a consistent and concise argumentation.

Notice that a key characteristic of the language in this performance level description is that it *describes* the work. Thus for students who aspire to this high level, the rubric depicts for them what their work needs to look like in order to reach that goal.

In contrast, if performance level descriptions are written in evaluative language (for example, if the performance level description above had read, "The paper shows excellent manuscript structure"), the rubric does not give students the information they need to further their learning. Rubrics written in evaluative language do not give students a depiction of work

at that level and, therefore, do not provide a clear description of the learning goal. An example of evaluative language used in a rubric can be found in the performance level descriptions for one of the criteria of an oral communication rubric (Avanzino, 2010, p. 109). This is the performance level description for Level 2 (Adequate) on the criterion of Delivery:

Speaker's delivery style/use of notes (manuscript or extemporaneous) is average; inconsistent focus on audience.

Notice that the key word in the first part of the performance level description, "average," does not give any information to the student about what average delivery looks like in regard to style and use of notes. The second part of the performance level description, "inconsistent focus on audience," is descriptive and gives students information about what Level 2 performance looks like in regard to audience focus.

## RESULTS AND DISCUSSION

The 46 studies yielded 51 different rubrics because several studies included more than one rubric. The two sections below take up results for each research question in turn.

### Type and Quality of Rubrics

Table 1 displays counts of the type and quality of rubrics found in the studies. Most of the rubrics (29 out of 51, 57%) were analytic, descriptive rubrics. This means they considered the criteria separately, requiring a separate decision about work quality for each criterion. In addition, it means that the performance

**TABLE 3 |** Validity evidence for rubrics.

| Study                                    | Level                   | Rubric topic & description   | Sample  | Validity evidence  |
|--|-------------------------|--|---|--|
| Avanzino, 2010                           | Undergraduate           | Oral communication<br>Analytic rubric with 3 criteria, 3 levels with mostly descriptive plds   | 230 speeches (112 individual, 118 group)  | Based on student learning outcomes;<br>Subject expert review   |
| Bauer and Cole, 2012                     | Undergraduate           | Chemistry guided-inquiry activities Rating scale, 0-3, on 15 indicators of POGIL (process oriented guided inquiry learning)  | 60 science faculty, 4 manipulated versions of the task  | Rubric was sensitive enough to distinguish four versions of the activity   |
| Britton et al., 2017                     | Undergraduate           | Team-Q Rubric for individual teamwork skills<br>Final version: 5 criteria, each with behavioral descriptions, rated with a 5-level frequency scale (never to always)   | 70 students in a theater history and literature course, 24 of whom gave full consent            | Factor analysis yielded a one-factor solution  |
| Chasteen et al., 2012                    | Undergraduate           | Physics, electromagnetism<br>Detailed task-specific point schemes for each task  | 103 students in 3 courses (final version), 432 students in 14 courses during test development   | Expert feedback;<br>Student interviews[<br>Student results differed by course (could differentiate types of instruction),<br>criterion-related evidence (to physics grades)  |
| Cho et al., 2006                         | Undergraduate, graduate | Writing<br>Analytic rubric with 3 criteria and 7 levels; Plds were somewhat descriptive but relied on counting (e.g., "all but one argument...") or rating-scale language  | 708 students in 16 courses over 3 years from 4 universities                                     | Correlations of student ratings with instructor and expert ratings   |
| Ciorba and Smith, 2009                   | Undergraduate           | Music – Instrumental and vocal performance<br>Analytic rubric with 3 criteria and descriptive plds at 5 levels   | 28 panels of judges, 359 music students' performances   | Scores rose by year (Fr-Soph-Jr-Sr);<br>Scale intercorrelations (internal validity evidence)   |
| Garcia-Ros, 2011                         | Undergraduate           | Oral presentation<br>14 criteria organized into 4 areas. 4 levels (0–3) with descriptive plds  | 64 educational psychology students  | Students' perceptions  |
| Hancock and Brundage, 2010               | Graduate                | Graduate Student Development Profile for Speech-Language Pathology students  | Pilot 26 first year students, then applied whole-program  | Demonstrated student growth over time;<br>Faculty perceptions  |
| Jonsson, 2014                            | Graduate                | 3 rubrics<br>Survey construction rubric in epidemiology: analytic, general rubric, 2 criteria, 4 levels with plds for each<br>House inspection rubric in real estate program: more like a checklist, w/ multiple criteria and a tally of facts and reasoning for each<br>Patient communication rubric in dental program: indicators for each of several criteria | 13 statistics students in an epidemiology program, 105 real estate students, 48 dental students | Students found the rubrics transparent and useful. Criteria were aligned with assignments, "thereby inviting the students to use the rubrics as guides to performance, as well as tools for self-assessment and reflection" (p. 849).<br>Results were interpreted to mean that rubrics made assessment expectations explicit for students. |
| Kocakülah, 2010                          | Undergraduate           | Physics – Newton's Laws of Motion problems<br>Rubric style point scheme; Analytic rubric with 6 criteria and descriptive plds at 5 levels, but points vary depending on the criterion  | 153 physics students in 4 classes   | Students' mean peer scores were same as Instructor scores  |
| Latifa et al., 2015                      | Undergraduate           | Practical Rating Rubric of Speaking Test<br>Holistic grading rubric with 5 levels (0-4), 5 criteria, mostly counting (e.g., percentage of errors)  | 12 English speaking lecturers in several institutions in Indonesia                              | Lecturers found the grading scale easy to use. Authors asserted they compared it with analytic scoring.  |
| Menéndez-Varela and Gregori-Giralt, 2016 | Undergraduate           | Service learning projects<br>2 analytic rubrics. Content: 4 criteria, 4 levels each, w/ descriptive plds. Oral presentation: 5 criteria, 4 levels, descriptive plds except for time  | 84 history of art students  | Three factors: Project content, Oral presentation skills, and Difficulty   |
| Moni et al., 2005                        | Undergraduate           | Concept maps – Physiology<br>Study was done using original "rubric," which was a point scheme for the concept map task. Revised rubric was an analytic rubric, 3 criteria, 5 levels, descriptive plds, based on student & faculty feedback   | 62 students, 2 faculty (plus 1 faculty advisor)   | Student perceptions;<br>Faculty perceptions  |
| Pagano et al., 2008                      | Undergraduate           | Writing (College composition)<br>Analytic rubric, 6 levels with descriptive plds at 3 of the levels (1-2, 3-4, 5-6)  | 6 institutions year 1, 5 institutions year 2  | Scores increased from early to late in the semester  |

(Continued)

TABLE 3 | Continued

| Study                   | Level         | Rubric topic & description   | Sample  | Validity evidence   |
|-------------------------|---------------|--|---|---|
| Prins et al., 2016      | Undergraduate | Research theses in education<br>Analytic rubric, 6 criteria, 3 levels, descriptive plds for levels 2 "must have" and 3 "nice to have" (where 1 was assumed to be "does not have")  | 105 students  | Studied student use and perceptions via questionnaire. Students felt rubrics had 4 functions (based on a factor analysis of questionnaire). Students who got lower grades on the task reported beginning to apply the rubric's criteria later. Faculty wanted another level to distinguish good from excellent work.  |
| Reddy, 2011             | Graduate      | Business Cases, Business Projects<br>Business case study rubric (4 dim); business project rubric (7 dim), each with descriptive plds at 4 levels   | 35 instructors, 95 business students, 2 institutions                                    | Expert review;<br>Student perceptions   |
| Rezaei and Lovorn, 2010 | Graduate      | Writing<br>Analytic rubrics with 5 criteria and descriptive plds at 4 levels; descriptions somewhat inferential (e.g., "limited understanding")  | 467 graduate students   | Quasi-experiment investigating influence of construct-irrelevant factors  |
| Schreiber et al., 2012  | Undergraduate | Public Speaking Competence Rubric<br>Analytic rubric with 9 criteria & 2 optional criteria, with descriptive plds at 5 levels  | Study 1, 5 coders, 45 speeches; Study 2, 3 undergraduate + 1 faculty coder, 50 speeches | Factor analysis (internal structure evidence); Criterion-related evidence (correlation of rubric scores for speeches with grades assigned to the speeches using different scoring schemes during the semester)  |
| Stellmack et al., 2009  | Undergraduate | Writing APA-style introductions<br>Analytic rubric with 8 dimensions with descriptive plds at 4 levels   | 40 papers, 3 researcher/graders   | Criterion-related evidence (Spearman correlation with independent judge)  |
| Timmerman et al., 2011  | Undergraduate | Science writing Analytic rubric with 15 criteria and descriptive plds at 4 levels  | 142 lab reports, 9 trained and 8 'natural' graduate student raters                      | Grader (graduate student) perceptions; Faculty (expert) review  |
| Urios et al., 2015      | Undergraduate | Teamwork and oral & written communication skills, in a chemical engineering degree<br>3 main criteria and subcriteria, with rating-scale language in 2 to 4 levels under each, mostly about surface features   | 2 groups, 30 students in each, 1 teacher & teaching assistant in each                   | Validation questionnaire. Students lacked knowledge of the use of rubrics, lacked adaptability and were somewhat resistant. Also "lack of commitment and proactivity in the teaching/learning process" p. 147.  |
| Wald et al., 2012       | Graduate      | Reflective writing<br>Analytic rubric with 5 criteria (+1 optional) and descriptive plds at 4 levels   | 10–60 narratives over 5 trials  | Rubric content based on literature  |
| Wallace et al., 2011    | Undergraduate | Astronomy – Cosmology<br>Task-specific, holistic rubrics for each test item, with 5 levels   | 65 responses from 21 students, 9 items  | Rubric content based on student responses to tasks  |
| Young, 2013             | Undergraduate | Physiotherapy clinical demonstrations<br>Holistic proforma used mostly rating-scale language, 5 levels, with some highly inferential description, 1/2 page; Analytic rubric was very complicated, more of a point scheme, 5 criteria (+safety pass/fail), 5 levels-to rate that required counting behaviors listed from the standards, 3 pages | 67 students   | Students' self-efficacy to grade was greater for the proforma than the rubric. Students felt rubric aided evaluation more than proforma at first (when they needed the behaviors listed explicitly) but changed in perception of competence to use the proforma by the end of the semester. Rubric was more useful for learning, but proforma was easier to use to score. |

plds, Performance Level Descriptions.

level descriptions used descriptive, as opposed to evaluative, language, which is expected to be more supportive of learning. Most commonly, these rubrics described four (14) or five (8) performance levels.

Four of the 51 rubrics (8%) were holistic, descriptive rubrics. This means they considered the criteria simultaneously, requiring one decision about work quality across all criteria at once. In addition, the performance level descriptions used the desired descriptive language.

Three of the rubrics were descriptive and task-specific. One of these was an analytic rubric and two were holistic rubrics.

None of the three could be shared with students, because they would "give away" answers. Such rubrics are more useful for grading than for formative assessment supporting learning. This does not necessarily mean the rubrics were not of quality, because they served well the grading function for which they were designed. However, they represent a missed opportunity to support learning as well as grading.

A few of the rubrics were not written in a descriptive manner. Six of the analytic rubrics and one of the holistic rubrics used rating scale language and/or listed counts of occurrences of elements in the work, instead of describing the quality of student

**TABLE 4 |** Descriptive case studies about developing and using rubrics.

| Study                    | Level         | Rubric topic & description   | Sample  |
|--------------------------|---------------|--|---|
| Bissell and Lemons, 2006 | Undergraduate | Introductory Biology Paper-and-Pencil Tasks<br>Detailed task-specific point schemes for grading biology paper-and-pencil tasks   | 150 students in 1 introductory biology course   |
| Bowen, 2017              | Undergraduate | Visual Literacy Competency Holistic rubric with 5 levels based on the SOLO taxonomy  | 2 courses, popular culture & visual rhetoric; applied rubric to 1 assignment in each course |
| Davidowitz et al., 2005  | Undergraduate | Rubric for flow diagrams in chemistry labs<br>Analytic rubric with plds using mostly rating-scale language (some descriptive) in 4 levels  | 133 flow diagrams from 16 students  |
| Dinur and Sherman, 2009  | Undergraduate | Business Case Study Presentation<br>3 rubrics, 2 of which were true rubrics. Content rubric was a 1–5 rating scale on 9 criteria; Oral presentation rubric was an analytic rubric with plds using frequency-scale language on 4 levels of 4 criteria; Written Assignment rubrics was an analytic rubric with 8 criteria (only 1 of which was about content) and descriptive plds at 4 levels | 159 business students   |
| Fraser et al., 2005      | Undergraduate | Business Writing<br>Analytic rubric with 6 criteria and descriptive plds at 5 levels   | Results summarized, sample size not given   |
| Knight, 2006             | Undergraduate | Information Literacy (Annotated Bibliographies)<br>Analytic rubric with 5 criteria and descriptive plds at 3 levels, but the descriptions include a lot of counting elements   | 260 bibliographies with 10 citations in each  |

*plds, Performance Level Descriptions.*

learning and performance. Thus 7 out of 51 (14%) of the rubrics were not of the quality that is expected to be best for student learning (Arter and McTighe, 2001; Arter and Chappuis, 2006; Andrade, 2010; Brookhart, 2013).

Finally, eight of the 51 rubrics (16%) were not rubrics but rather rating scales (5) or point schemes for grading (3). It is possible that the authors were not aware of the more nuanced meaning of “rubric” currently used by educators and used the term in a more generic way to mean any scoring scheme.

As the heart of Research Question 1 was about the potential of the rubrics used to contribute to student learning, I also coded the studies according to whether the rubrics were used with students or whether they were just used by instructors for grading. Of the 46 studies, 26 (56%) reported using the rubrics with students and 20 (43%) did not use rubrics with students but rather used them only for grading.

## Relation of Rubric Type to Reliability, Validity, and Learning

Different studies reported different characteristics of their rubrics. I charted studies that reported evidence for the reliability of information from rubrics (Table 2) and the validity of information from rubrics (Table 3). For the sake of completeness, Table 4 lists six studies that presented their work with rubrics in a descriptive case-study style that did not fit easily into Table 2 or Table 3 or in Table 5 (below) about the effects of rubrics on learning. With the inclusion of Table 4, readers have descriptions of all 51 rubrics in all 46 studies reported under Research Question 1.

Reliability was most commonly studied as inter-rater reliability, arguably the most important for rubrics because judgment is involved in matching student work with

performance level descriptions, or as internal consistency among criteria. Construct validity was addressed with a variety of methods, from expert review to factor analysis; some studies also addressed consequential evidence for validity with student or faculty questionnaires. No discernable patterns were found that indicated one form of rubric was preferable to another in regard to reliability or validity. Although this conforms to my hypothesis, this result is also partly because most of the studies’ reported results and experience with rubrics were positive, no matter what type of rubric was used.

Table 5 describes 13 studies of the effects of rubrics on learning or motivation, all with positive results. Learning was most commonly operationalized as improvement in student work. Motivation was typically operationalized as student responses to questionnaires. In these studies as well, no discernable pattern was found regarding type of rubric. Despite the logical and learning-based arguments made in the literature and summarized in the introduction to this article, rubrics with both descriptive and evaluative performance level descriptions both led to at least some positive results for students. Eight of these studies used descriptive rubrics and five used evaluative rubrics. It is possible that the lack of association of type of rubric with study findings is a result of publication bias, because most of the studies had good things to say about rubrics and their effects. The small sample size (13 studies) may also be an issue.

## CONCLUSIONS

Rubrics are becoming more and more evident as part of assessment in higher education. Evidence for that claim is simply the number of studies that are published investigating this new and growing interest and the assertions made in those studies about rising interest in rubrics.



**TABLE 5 |** Studies of the effects of rubric use on student learning and motivation to learn.

| Study                                    | Level                    | Rubric topic & description  | Sample   | Design                 | Findings  |
|--|--------------------------|---|--|------------------------|---|
| Andrade and Du, 2005                     | Undergraduate            | Educational Psychology Learning Vignettes Performance Rubric, Analytic rubric with 6 weighted criteria and descriptive plds at 4 levels   | 14 teacher education students who had used rubrics in Ed Psych                       | Focus groups           | Students used rubrics to determine teacher's expectations, plan production, check their work in progress, and guide and reflect on feedback. Some students only checked the A and B levels of the rubric, and some saw rubrics as a way to "give teachers what they want."  |
| Ash et al., 2005                         | Undergraduate            | Service learning objectives, Critical thinking<br>Holistic rubric for service learning objectives, listed according to level of thinking, 0–4 (0 not described), so the learning objectives formed the descriptions; Holistic critical thinking rubric, 4 levels, 8 simultaneous criteria, descriptive plds | 14 students in 2 classes   | Pre-experimental       | Improvement across drafts was noted, with the Academic criterion being the most difficult for students. Improvement in first drafts across the semester was also noted, but smaller, and again the Academic criterion was the hardest.  |
| Britton et al., 2017                     | Undergraduate            | Team-Q Rubric for individual teamwork skills<br>Final version: 5 criteria, each with behavioral descriptions, rated with a 5-level frequency scale (never to always)  | 70 students in a theater history and literature course, 24 of whom gave full consent | Instrument development | Significant improvement in teamwork skills from first time to second time in both self-ratings and peer ratings. External ratings improved from Time 1 to Time 2 but not significantly so.  |
| Howell, 2011                             | Undergraduate            | Juvenile delinquency course assignment rubric<br>Holistic grading rubric, somewhat task-specific, plds for each of 4 levels, which were then converted to points for grading  | 80 students in 2 sections of the instructor's own course                             | Quasi-experimental     | Controlling for college year, criminal justice major (vs. not), pretest score and gender, being in the treatment group (having rubrics provided with the assignment) predicted achievement ( $\beta = 0.488$ ). The only other large predictor was college year. Student achievement was higher when rubrics were used. |
| Howell, 2014                             | Undergraduate            | Juvenile delinquency course assignment rubric<br>Holistic grading rubric, somewhat task-specific, plds for each of 4 levels, which were then converted to points for grading  | 76 students in 2 sections of the instructor's own course                             | Quasi-experimental     | Treatment group (completed an assignment using a grading rubric) scored higher than comparison group (same assignment, no rubric). Regression showed rubric used contributed significantly after controlling for baseline course knowledge and gpa.   |
| Kerby and Romine, 2010                   | Undergraduate & graduate | Oral communications and presentation<br>Analytic rubric with 8 criteria and descriptive plds at 3 levels  | 1 business accounting program  | Case study             | Oral presentation skills improved from sophomore to senior years, did not further improve in graduate level, which the researchers attributed to more complex material to present.  |
| Kocakulah, 2010                          | Undergraduate            | Newton's Laws of Motion problem solving<br>Rubric style point scheme; Analytic rubric with 6 criteria and descriptive plds at 5 levels, but points vary depending on the criterion  | 153 physics students in 4 classes  | Quasi-experimental     | Students who took part in the designing and using of a rubric, performed better in solving problems than those who had the same instruction but no rubric.  |
| McCormick et al., 2007                   | Undergraduate            | Self-assessment of Executive Leadership<br>Analytic rubric with 6 criteria and 8 levels (0–7), with descriptive plds at levels 2, 4, and 6  | 44 seniors in a leadership education course  | Pre-experimental       | Student perceived competence increased over the semester. Half of the students accurately estimated their competence (based on final exam), the other half underestimated their competence.   |
| Menéndez-Varela and Gregori-Giralt, 2016 | Undergraduate            | Service learning projects<br>2 analytic rubrics. Content: 4 criteria, 4 levels each, w/descriptive plds. Oral presentation: 5 criteria, 4 levels, descriptive plds except for time  | 84 history of art students   | Validity study         | Significant increase in scores (quality of projects) from stage 1 to stage 3 of the study, overall and for each of 5 raters individually; work quality increased as rubric use was repeated   |
| Petkov and Petkova, 2006                 | Undergraduate            | Business Projects<br>13 criteria grouped into 4 areas, with rating-scale language at 4 levels   | 20 students fall (rubric), 20 students spring (no rubric)                            | Pre-experimental       | Rubrics group achievement was higher than the comparison group.   |

(Continued)

TABLE 5 | Continued

| Study                   | Level         | Rubric topic & description  | Sample  | Design                   | Findings   |
|-------------------------|---------------|---|---|--------------------------|--|
| Reynolds-Keefer, 2010   | Undergraduate | Writing<br>Analytic rubric with 5 criteria and descriptive plds for 6 levels  | 45 ed psych students  | Open-ended questionnaire | Pre-service teachers who used rubrics as students reported being more likely to use rubrics in their own teaching.   |
| Ritchie, 2016           | Undergraduate | Oral presentations in biology<br>"Rubric" was really a rating scale with 15 criteria org under "content," organization, & delivery, scored 1–5, "poor/absent" to "no change needed" | 39 students in 2 sections (1 w/rubric self-assessment & 1 without); each gave 2 presentations | Pre-experimental         | Students in self-assessment w/rubrics group improved more in 2nd presentation, with less variability. All viewed their videotaped presentation (cf. 47% of control grp). Peer assessment accurate (compared with instructor), self-assessment was not. |
| Vandenberg et al., 2010 | Undergraduate | Financial analysis project<br>Analytic rubric with 5 criteria and descriptive plds for 5 levels   | 49 students in 3 sections of the course   | Pre-experimental         | Students who used rubrics scored significantly higher on two of three sections of the project. Students with rubrics felt the requirements of the assignment were more clearly communicated than those without.  |

*Plds, Performance Level Descriptions.*

Research Question 1 asked about the type and quality of rubrics published in studies of rubrics in higher education. The number of criteria varies widely depending on the rubric and its purpose. Three, four, and five are the most common number of levels. While most of the rubrics are descriptive—the type of rubrics generally expected to be most useful for learning—many are not. Perhaps most surprising, and potentially troubling, is that only 56% of the studies reported using rubrics with students. If all that is required is a grading scheme, traditional point schemes or rating scales are easier for instructors to use. The value of a rubric lies in its formative potential (Panadero and Jonsson, 2013), where the same tool that students can use to learn and monitor their learning is then used for grading and final evaluation by instructors.

Research Question 2 asked whether rubric type and quality were related to measurement quality (reliability and validity) or effects on learning and motivation to learn. Among studies in this review, reported reliability and validity was not related to type of rubric. Reported effects on learning and/or motivation were not related to type of rubric. The discussion above speculated that part of the reason for these findings might be publication bias, because only studies with good effects—whatever the type of rubric they used—were reported.

However, we should not dismiss all the results with a hand-wave about publication bias. All of the tools in the studies of rubrics—true rubrics, rating scales, checklists—had criteria. The differences were in the type of scale and scale descriptions used. Criteria lay out for students and instructors what is expected in student work and, by extension, what it looks like when evidence of intended learning has been produced. Several of the articles stated explicitly that the point of rubrics was to make assignment expectations explicit (e.g., Andrade and Du, 2005; Fraser et al., 2005; Reynolds-Keefer, 2010; Vandenberg et al., 2010; Jonsson, 2014; Prins et al., 2016). The criteria are the assignment expectations: the qualities the final work should display. The performance level descriptions instantiate those expectations at different levels of competence. Thus, one firm conclusion from this review is that appropriate criteria are the

key to effective rubrics. Trivial or surface-level criteria will not draw learning goals for students as clearly as substantive criteria. Students will try to produce what is expected of them. If the criterion is simply having or counting something in their work (e.g., "has 5 paragraphs"), students need not pay attention to the quality of what their work has. If the criterion is substantive (e.g., "states a compelling thesis"), attention to quality becomes part of the work.

It is likely that appropriate performance level descriptions are also key for effective rubrics, but this review did not establish this fact. A major recommendation for future research is to design studies that investigate how students use the performance level descriptions as they work, in monitoring their work, and in their self-assessment judgments. Future research might also focus on two additional characteristics of rubrics (Dawson, 2017): users and uses and judgment complexity. Several studies in this review established that students use rubrics to make expectations explicit. However, in only 56% of the studies were rubrics used with students, thus missing the opportunity to take advantage of this important rubric function. Therefore, it seems important to seek additional understanding of users and uses of rubrics. In this review, judgment complexity was a clear issue for one study (Young, 2013). In that study, a complex rubric was found more useful for learning, but a holistic rating scale was easier to use once the learning had occurred. This hint from one study suggests that different degrees of judgment complexity might be more useful in different stages of learning.

Rubrics are one way to make learning expectations explicit for learners. Appropriate criteria are key. More research is needed that establishes how performance level descriptions function during learning and, more generally, *how* students use rubrics for learning, not just *that* they do.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## REFERENCES

- Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership* 57, 13–18. Available online at: <http://www.ascd.org/publications/educational-leadership/feb00/vol57/num05/Using-Rubrics-to-Promote-Thinking-and-Learning.aspx>
- Andrade, H., and Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Pract. Assess. Res. Eval.* 10, 1–11. Available online at: <http://pareonline.net/pdf/v10n3.pdf>
- Andrade, H., and Heritage, M. (2017). *Using Assessment to Enhance Learning, Achievement, and Academic Self-Regulation*. New York, NY: Routledge.
- Andrade, H. L. (2010). “Students as the definitive source of formative assessment: academic self-assessment and the self-regulation of learning,” in *Handbook of Formative Assessment*, eds H. L. Andrade and G. J. Cizek (New York, NY: Routledge), 90–105.
- Arter, J. A., and Chappuis, J. (2006). *Creating and Recognizing Quality Rubrics*. Boston: Pearson.
- Arter, J. A., and McTighe, J. (2001). *Scoring Rubrics in the Classroom: Using Performance Criteria for Assessing and Improving Student Performance*. Thousand Oaks, CA: Corwin.
- Ash, S. L., Clayton, P. H., and Atkinson, M. P. (2005). Integrating reflection and assessment to capture and improve student learning. *Mich. J. Comm. Serv. Learn.* 11, 49–60. Available online at: <http://hdl.handle.net/2027/spo.3239521.0011.204>
- Avanzino, S. (2010). Starting from scratch and getting somewhere: assessment of oral communication proficiency in general education across lower and upper division courses. *Commun. Teach.* 24, 91–110. doi: 10.1080/17404621003680898
- Bauer, C. F., and Cole, R. (2012). Validation of an assessment rubric via controlled modification of a classroom activity. *J. Chem. Educ.* 89, 1104–1108. doi: 10.1021/ed2003324
- Bell, A., Mladenovic, R., and Price, M. (2013). Students’ perceptions of the usefulness of marking guides, grade descriptors and annotated exemplars. *Assess. Eval. High. Educ.* 38, 769–788. doi: 10.1080/02602938.2012.714738
- Bissell, A. N., and Lemons, P. R. (2006). A new method for assessing critical thinking in the classroom. *BioScience*, 56, 66–72. doi: 10.1641/0006-3568(2006)056[0066:ANMFAC]2.0.CO;2
- Bowen, T. (2017). Assessing visual literacy: a case study of developing a rubric for identifying and applying criteria to undergraduate student learning. *Teach. High. Educ.* 22, 705–719. doi: 10.1080/13562517.2017.1289507
- Britton, E., Simper, N., Leger, A., and Stephenson, J. (2017). Assessing teamwork in undergraduate education: a measurement tool to evaluate individual teamwork skills. *Assess. Eval. High. Educ.* 42, 378–397. doi: 10.1080/02602938.2015.1116497
- Brookhart, S. M. (2013). *How to Create and Use Rubrics for Formative Assessment and Grading*. Alexandria, VA: ASCD.
- Brookhart, S. M., and Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educ. Rev.* 67, 343–368. doi: 10.1080/00131911.2014.929565
- Brookhart, S. M., and Nitko, A. J. (2019). *Educational Assessment of Students, 8th Edn*. Boston, MA: Pearson.
- Chasteen, S. V., Pepper, R. E., Caballero, M. D., Pollock, S. J., and Perkins, K. K. (2012). Colorado Upper-Division Electrostatics diagnostic: a conceptual assessment for the junior level. *Phys. Rev. Spec. Top. Phys. Educ. Res.* 8:020108. doi: 10.1103/PhysRevSTPER.8.020108
- Cho, K., Schunn, C. D., and Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *J. Educ. Psychol.* 98, 891–901. doi: 10.1037/0022-0663.98.4.891
- Ciorba, C. R., and Smith, N. Y. (2009). Measurement of instrumental and vocal undergraduate performance juries using a multidimensional assessment rubric. *J. Res. Music Educ.* 57, 5–15. doi: 10.1177/0022429409333405
- Davidowitz, B., Rollnick, M., and Fakudze, C. (2005). Development and application of a rubric for analysis of novice students’ laboratory flow diagrams. *Int. J. Sci. Educ.* 27, 43–59. doi: 10.1080/0950069042000243754
- 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
- DeWever, B., Van Keer, H., Schellens, T., and Valke, M. (2011). Assessing collaboration in a wiki: the reliability of university students’ peer assessment. *Internet High. Educ.* 14, 201–206. doi: 10.1016/j.iheduc.2011.07.003
- Dinur, A., and Sherman, H. (2009). Incorporating outcomes assessment and rubrics into case instruction. *J. Behav. Appl. Manag.* 10, 291–311.
- Facione, N. C., and Facione, P. A. (1996). Externalizing the critical thinking in knowledge development and clinical judgment. *Nurs. Outlook* 44, 129–136. doi: 10.1016/S0029-6554(06)80005-9
- Falchikov, N., and Boud, D. (1989). Student self-assessment in higher education: a meta-analysis. *Rev. Educ. Res.* 59, 395–430.
- Fraser, L., Harich, K., Norby, J., Brzovic, K., Rizkallah, T., and Loewy, D. (2005). Diagnostic and value-added assessment of business writing. *Bus. Commun. Q.* 68, 290–305. doi: 10.1177/1080569905279405
- Garcia-Ros, R. (2011). Analysis and validation of a rubric to assess oral presentation skills in university contexts. *Electr. J. Res. Educ. Psychol.* 9, 1043–1062.
- Hancock, A. B., and Brundage, S. B. (2010). Formative feedback, rubrics, and assessment of professional competency through a speech-language pathology graduate program. *J. All. Health*, 39, 110–119.
- Hattie, J., and Timperley, H. (2007). The power of feedback. *Rev. Educ. Res.* 77, 81–112. doi: 10.3102/003465430298487
- Howell, R. J. (2011). Exploring the impact of grading rubrics on academic performance: findings from a quasi-experimental, pre-post evaluation. *J. Excell. Coll. Teach.* 22, 31–49.
- Howell, R. J. (2014). Grading rubrics: hoopla or help? *Innov. Educ. Teach. Int.* 51, 400–410. doi: 10.1080/14703297.2013.785252
- Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assess. Eval. High. Educ.* 39, 840–852. doi: 10.1080/02602938.2013.875117
- 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
- Kerby, D., and Romine, J. (2010). Develop oral presentation skills through accounting curriculum design and course-embedded assessment. *Journal of Education for Business*, 85, 172–179. doi: 10.1080/08832320903252389
- Knight, L. A. (2006). Using rubrics to assess information literacy. *Ref. Serv. Rev.* 34, 43–55. doi: 10.1108/00907320610640752
- Kocakulah, M. (2010). Development and application of a rubric for evaluating students’ performance on Newton’s Laws of Motion. *J. Sci. Educ. Technol.* 19, 146–164. doi: 10.1007/s10956-009-9188-9
- Latifa, A., Rahman, A., Hamra, A., Jabu, B., and Nur, R. (2015). Developing a practical rating rubric of speaking test for university students of English in Parepare, Indonesia. *Engl. Lang. Teach.* 8, 166–177. doi: 10.5539/elt.v8n6p166
- Lewis, L. K., Stiller, K., and Hardy, F. (2008). A clinical assessment tool used for physiotherapy students—is it reliable? *Physiother. Theory Pract.* 24, 121–134. doi: 10.1080/09593980701508894
- McCormick, M. J., Dooley, K. E., Lindner, J. R., and Cummins, R. L. (2007). Perceived growth versus actual growth in executive leadership competencies: an application of the stair-step behaviorally anchored evaluation approach. *J. Agric. Educ.* 48, 23–35. doi: 10.5032/jae.2007.02023
- Menéndez-Varela, J., and Gregori-Giralt, E. (2016). The contribution of rubrics to the validity of performance assessment: a study of the conservation-restoration and design undergraduate degrees. *Assess. Eval. High. Educ.* 41, 228–244. doi: 10.1080/02602938.2014.998169
- Moni, R. W., Beswick, E., and Moni, K. B. (2005). Using student feedback to construct an assessment rubric for a concept map in physiology. *Adv. Physiol. Educ.* 29, 197–203. doi: 10.1152/advan.00066.2004
- Newman, L. R., Lown, B. A., Jones, R. N., Johansson, A., and Schwartzstein, R. M. (2009). Developing a peer assessment of lecturing instrument: lessons learned. *Acad. Med.* 84, 1104–1110. doi: 10.1097/ACM.0b013e3181ad18f9
- Nicholson, P., Gillis, S., and Dunning, A. M. (2009). The use of scoring rubrics to determine clinical performance in the operating suite. *Nurse Educ. Today* 29, 73–82. doi: 10.1016/j.nedt.2008.06.011
- Nordrum, L., Evans, K., and Gustafsson, M. (2013). Comparing student learning experiences of in-text commentary and rubric-articulated feedback: strategies for formative assessment. *Assess. Eval. High. Educ.* 38, 919–940. doi: 10.1080/02602938.2012.758229

- Pagano, N., Bernhardt, S. A., Reynolds, D., Williams, M., and McCurrie, M. (2008). An inter-institutional model for college writing assessment. *Coll. Composition Commun.* 60, 285–320.
- Panadero, E., and Jonsson, 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
- Petkov, D., and Petkova, O. (2006). Development of scoring rubrics for IS projects as an assessment tool. *Issues Informing Sci. Inform. Technol.* 3, 499–510. doi: 10.28945/910
- Prins, F. J., de Kleijn, R., and van Tartwijk, J. (2016). Students' use of a rubric for research theses. *Assess. Eval. High. Educ.* 42, 128–150. doi: 10.1080/02602938.2015.1085954
- Reddy, M. Y. (2011). Design and development of rubrics to improve assessment outcomes: a pilot study in a master's level Business program in India. *Qual. Assur. Educ.* 19, 84–104. doi: 10.1108/09684881111107771
- Reddy, Y., and Andrade, H. (2010). A review of rubric use in higher education. *Assess. Eval. High. Educ.* 35, 435–448. doi: 10.1080/02602930902862859
- Reynolds-Keefer, L. (2010). Rubric-referenced assessment in teacher preparation: an opportunity to learn by using. *Pract. Assess. Res. Eval.* 15, 1–9. Available online at: <http://pareonline.net/getvn.asp?v=15&n=8>
- Rezaei, A., and Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assess. Writing*, 15, 18–39. doi: 10.1016/j.asw.2010.01.003
- Ritchie, S. M. (2016). Self-assessment of video-recorded presentations: does it improve skills? *Act. Learn. High. Educ.* 17, 207–221. doi: 10.1177/1469787416654807
- Rochford, L., and Borchert, P. S. (2011). Assessing higher level learning: developing rubrics for case analysis. *J. Educ. Bus.* 86, 258–265. doi: 10.1080/08832323.2010.512319
- Sadler, D. R. (2014). The futility of attempting to codify academic achievement standards. *High. Educ.* 67, 273–288. doi: 10.1007/s10734-013-9649-1
- Schamber, J. F., and Mahoney, S. L. (2006). Assessing and improving the quality of group critical thinking exhibited in the final projects of collaborative learning groups. *J. Gen. Educ.* 55, 103–137. doi: 10.1353/jge.2006.0025
- Schreiber, L. M., Paul, G. D., and Shibley, L. R. (2012). The development and test of the public speaking competence rubric. *Commun. Educ.* 61, 205–233. doi: 10.1080/03634523.2012.670709
- Stellmack, M. A., Konheim-Kalkstein, Y. L., Manor, J. E., Massey, A. R., and Schmitz, J. P. (2009). An assessment of reliability and validity of a rubric for grading APA-style introductions. *Teach. Psychol.* 36, 102–107. doi: 10.1080/00986280902739776
- Timmerman, B. E. C., Strickland, D. C., Johnson, R. L., and Payne, J. R. (2011). Development of a 'universal' rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assess. Eval. High. Educ.* 36, 509–547. doi: 10.1080/02602930903540991
- 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.* 14, 281–294. doi: 10.1080/09695940701591867
- Urios, M. I., Rangel, E. R., Tomàs, R. B., Salvador, J. T., Garcíá, F. C., and Piquer, C. F. (2015). Generic skills development and learning/assessment process: use of rubrics and student validation. *J. Technol. Sci. Educ.* 5, 107–121. doi: 10.3926/jotse.147
- Vandenberg, A., Stollak, M., McKeag, L., and Obermann, D. (2010). GPS in the classroom: using rubrics to increase student achievement. *Res. High. Educ. J.* 9, 1–10. Available online at: <http://www.aabri.com/manuscripts/10522.pdf>
- Wald, H. S., Borkan, J. M., Taylor, J. S., Anthony, D., and Reis, S. P. (2012). Fostering and evaluating reflective capacity in medical education: developing the REFLECT rubric for assessing reflective writing. *Acad. Med.* 87, 41–50. doi: 10.1097/ACM.0b013e31823b55fa
- Wallace, C. S., Prather, E. E., and Duncan, D. K. (2011). A study of general education Astronomy students' understandings of cosmology. Part II. Evaluating four conceptual cosmology surveys: a classical test theory approach. *Astron. Educ. Rev.* 10:010107. doi: 10.3847/AER2011030
- Young, C. (2013). Initiating self-assessment strategies in novice physiotherapy students: a method case study. *Assess. Eval. High. Educ.* 38, 998–1011. doi: 10.1080/02602938.2013.771255

**Conflict of Interest Statement:** The author declares 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 Brookhart. 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 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.





# Transparency or Stimulating Meaningfulness and Self-Regulation? A Case Study About a Programmatic Approach to Transparency of Assessment Criteria

Liesbeth K. J. Baartman<sup>1\*</sup> and Frans J. Prins<sup>2</sup>

<sup>1</sup> Research Group Vocational Education, Research Centre for Learning and Innovation, Utrecht University of Applied Sciences, Utrecht, Netherlands, <sup>2</sup> Department of Education, Faculty of Social and Behavioural Sciences, Utrecht University, Utrecht, Netherlands

## OPEN ACCESS

### Edited by:

Bronwen Cowie,  
University of Waikato, New Zealand

### Reviewed by:

Carmen Tomas,  
University of Nottingham,  
United Kingdom  
Jill Willis,  
Queensland University of Technology,  
Australia

### \*Correspondence:

Liesbeth K. J. Baartman  
liesbeth.baartman@hu.nl

### Specialty section:

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

**Received:** 04 May 2018

**Accepted:** 12 November 2018

**Published:** 28 November 2018

### Citation:

Baartman LKJ and Prins FJ (2018)  
Transparency or Stimulating  
Meaningfulness and Self-Regulation?  
A Case Study About a Programmatic  
Approach to Transparency of  
Assessment Criteria.  
Front. Educ. 3:104.  
doi: 10.3389/feduc.2018.00104

This exploratory case study focused on fostering meaning making of assessment criteria and standards at the module level and the course/programme level (the entire study plan), and the role of self-regulation in this meaning making process. The research questions that guided this study are: (1) How can students' meaning making of assessment criteria at the module level be fostered, (2) How can students' meaning making of assessment criteria at the programme level be fostered, and (3) How can self-regulation contribute to students' meaning making process? We explored the design and implementation of a rather new Master's programme in The Netherlands: The Master's Expert Teacher of Vocational Education (METVE). Interviews with three developers, three teachers, and 10 students of the METVE were analyzed. For each research question, several themes were derived from the data. Results indicate that meaning making takes place at the module level by using holistic assessment criteria and evaluative experiences, which allow students to make choices within the boundaries set by the assessment criteria. Meaning making at the programme level is experienced as much more difficult by students as well as teachers. The design of the METVE programme fosters meaning making at the programme level, but METVE teachers also express difficulties supporting this. Finally, we found that students perceive self-regulation as something extra for which they don't have enough time. Self-regulation at the programme level was not explicitly addressed and supported in the METVE, which makes it more difficult for some students to steer their learning process toward the role they are aiming for in professional practice after completing the Master's programme.

**Keywords:** assessment, criteria, transparency, assessment programme, self-regulation

## INTRODUCTION

Higher education aims to build a foundation for professionals in later work settings and social settings. In higher education, the specification of learning outcomes, and standards (the attainment levels) may be desirable in terms of transparency, but an unintended consequence may be to portray to students the idea that learning outcomes are a given (something done to them) and that good work means to work toward criteria set by others (Boud and Falchikov, 2006). In professional

practice, however, no lists or rubrics exist describing what “good work” looks like. Professionals have to be able to form their own complex judgments of their work and that of others, often in collaboration with colleagues, partners, customers, clients, etcetera, in short with all stakeholders directly or indirectly involved in their work (cf., evaluative judgement; Boud et al., 2018). If the above pictures professional practice and what is expected from students in later work settings, what are the implications for assessment and specifically the transparency of assessment criteria? Assessment criteria are often shared with students to communicate expectations and stimulate student performance in the “intended” direction (i.e., most of the times intended by the teacher), mainly at the module level. Transparency of assessment criteria may make clear to students what is expected of them. On the other hand, it may produce students who are more dependent on their teachers and may weaken rather than strengthen the development of self-regulated learning and learner autonomy (Torrance, 2007). From a programmatic perspective, transparency of assessment criteria may prevent students from choosing their own learning goals, their learning tasks and modules and, consequently, prevent them from assembling their own learning path during the curriculum. In other words, transparency of assessment criteria may be detrimental for the development of students’ self-regulatory and lifelong learning skills. Self-regulation refers to self-generated thoughts, feelings, and behaviors that are oriented to attaining goals (Zimmerman, 2000, 2002), which may concern the task level, the module level, the programme level as well as a lifelong learning perspective.

In this contribution, we therefore work out the argument that transparency not necessarily means that students get an exact picture of what is expected of them, but we propose that transparency could instead be viewed as meaning making of assessment criteria, both at the module level and at the programme/curriculum level. We add a curriculum level perspective to the discussion about transparency of assessment criteria, focusing on what is expected of students during the entire curriculum, at the end of the curriculum, and in later working life. In a case study, we explored how students’ long-term development throughout the curriculum can be brought to the forefront and how students’ meaning making of assessment criteria and self-regulatory skills may be stimulated.

The research questions that guided this study are: (1) How can students’ meaning making of assessment criteria at the module level be fostered, (2) How can students’ meaning making of assessment criteria at the programme level be fostered, and (3) How can self-regulation at the module level and programme level contribute to students’ meaning making process? In the remainder of this contribution, we first take the perspective of the module level. Then we shift to the programme level, focusing on students’ long-term learning process toward the programme or graduate learning outcomes. Third, using the framework of Zimmerman (2000; 2002) on self-regulation, we explore the role of self-regulatory skills at the module and programme level in meaning making of assessment criteria. We end with a single exploratory case study to explore meaning-making

of assessment criteria in practice, with varying degrees of success.

## MEANING MAKING OF ASSESSMENT CRITERIA AT THE MODULE LEVEL

In drive for transparency, standards and criteria at the module level (e.g., for assignments or exams) are often made explicit through (long) lists of criteria, benchmarks, rubrics, etc. Several researchers (e.g., Black and Wiliam, 1998; Rust et al., 2003; Wiliam, 2011) argue for the importance of clarifying the intended learning outcomes, because low achievement can be caused by students not knowing or understanding what is expected. On the other hand, students express disappointment about the over-reliance on written criteria to deliver clarity about assessment criteria and the lack of opportunities to internalize standards (Nicol and Macfarlane-Dick, 2006). Assessors use both explicit and tacit knowledge about standards when assessing student work (Bloxxham and Campbell, 2010; Price et al., 2011). Bloxxham and Campbell (2010) and Hawe and Dixon (2014) showed that when teachers do not share tacitly held criteria with students, this can result in misalignment between the judgments made by students and those made by the teacher. Understanding tacit criteria in a (work) community of practice takes place through an active, shared process rather than a one-way communication of explicit criteria to students. This is also confirmed in a recent literature review on teachers’ formative assessment practices, which also showed the importance of an active role of students in explicating and understanding learning goals and assessment criteria (Gulikers and Baartman, 2017). Students thus need to be actively engaged to develop a conceptualization of what constitutes quality if they are to improve their work and reach higher levels of performance (Sadler, 2009).

The ability to assess a piece of work against contextually appropriate standards is at the heart of “evaluative judgment” (e.g., Boud et al., 2018; Panadero and Broadbent, 2018). Research into evaluative judgment also offers suggestions for pedagogical practices in the classroom to stimulate students’ evaluative judgment capacity, which fits nicely with our ideas about meaning making of assessment criteria. Panadero and Broadbent argue for the importance of peer assessment and self-assessment, as these activities enhance evaluative judgment capacity. Other strategies include the use of scaffolding tasks, rubrics and exemplars to clarify and discuss assessment criteria and expectations (e.g., Fluckiger et al., 2010; Conway, 2011). Students can be confronted with a wide variety of authentic works, from other students attempting the same task and/or authentic products from “the real world,” review these good and bad examples and distill success criteria together (Fluckiger et al., 2010; Willis, 2011; Hawe and Dixon, 2014). In higher education, Fluckiger et al. (2010) describe and evaluate four strategies aimed to involve students as partners in the assessment process, to develop a learning climate, and to help students use assessment results to change their learning tactics.

Altogether, in a meaning making process these activities stimulate students to discover that different responses to an

assessment task may all result in valid products that comply with the quality criteria fit for the task. Or as Conway (2011) explains about his history lessons: “if the success criteria are shared and the students understand both what they are working toward and why, then they can take a lot of responsibility and we can allow a lot of variety. (p.4).”

## MEANING MAKING OF ASSESSMENT CRITERIA AT THE PROGRAMME LEVEL

So far, our discussion focused on transparency and meaning making of assessment criteria at the module level, for single assignments or exams. However, the ultimate goal of curricula in higher education is to prepare students for working and social life, and lifelong learning. Assessment involves making judgments about quality and identifying appropriate standards and criteria for the task at hand. This is as necessary to lifelong learning as it is to any formal educational experience (Boud, 2000). What constitutes quality is not a matter of one specific assignment or piece of work, but we view quality as a generalized attribute that can take specific forms or meanings in different contexts. Higher education aims to prepare learners to undertake such judgmental activities and to identify whether their work meets whatever standards are appropriate for the task at hand. To do so, Bok et al. (2013a) and Bok et al. (2013b) focus on stimulating students’ feedback-seeking behavior during an entire assessment programme. Students seek feedback from various sources during their clinical clerkships, depending on personal and interpersonal factors such as the students’ goal-orientation (focused on learning or on keeping a positive self-image) and the anticipated costs and benefits of the feedback. This feedback seeking behavior and judgments of quality also authentically mirror the ways many quality appraisals are made in everyday and work contexts by professionals.

Consequently, when it comes to transparency and meaning making of assessment criteria, this is not only important at the module level, but also (and maybe even more) at the programme level. In the Netherlands, the context of this study, we observe a drive toward detailed module specifications and explicit assessment criteria, a development Hughes et al. (2015) and Jessop and Tomas (2017) also describe in the UK. In this contribution, we therefore add a programme-level perspective to the discussion about transparency. In programme-focused assessment (van der Vleuten et al., 2012; Bok et al., 2013a) an arrangement of assessment methods is deliberately designed across the entire curriculum, combined and planned to optimize both robust decisions about students (summative) and student learning (formative). Rather than focus on specific or isolated assessments at the module level, a programme perspective focuses on the holistic developmental goals of the programme as a whole (Rust et al., 2012; Hughes et al., 2015). It follows then that such assessment is integrative in nature, trying to bring together “data points” or sources of information about students’ development that represent—in varying ways—the key programme outcomes (PASS position paper, 2012). Formatively, assessment activities are viewed as information sources that

provide a constant and longitudinal flow of information about student learning (Heeneman et al., 2015). The balance is shifted from summative to formative assessment to encourage students to think about longer term development rather than short term grade acquisition (Heeneman et al., 2015). An important starting point for programme-focused assessment is an overarching structure: the specification of the programme or graduate learning outcomes (Lokhoff et al., 2010; Hartley and Whitfield, 2011) and a number of levels or stepping stones that describe the development process toward these programme learning outcomes. These stepping stones are comparable to the learning progressions mentioned in Gulikers and Baartman’s review (2017) on teachers’ formative assessment practices. Key to stepping stones or learning progressions is that these specifications enable teachers and students to monitor progress on a longer term.

Programme learning outcomes are necessarily described in a holistic way as they need to capture the diversity of the (future) professional work context. Some concepts—like what constitutes a “good” or “tasty” dish—are in principle beyond the reach of formal definition. Experts in a professional domain can give valid and elaborate descriptions of what quality looks like in a particular specific instance (e.g., a cook can distinguish a good from a bad dish), but they are unable to do so for general cases. Sadler (2009) therefore argues for the use of holistic assessment criteria, because students need to be induced into judging what quality entails, without being bound by tightly specified criteria. Analytic grading constraints the scope of student work (one solution) and offers little imperative to explore alternative ways forward. The discussion between holistic and analytic or task-specific criteria is a complicated one, especially when it comes to meaning making of assessment criteria at the programme level. Previous research shows the advantages of task-specific criteria (Weigle, 2002; Jonsson and Svingby, 2007), as these criteria provide clear directions to students about what is expected. Govaerts et al. (2005) provide a more nuanced picture indicating that starting students prefer more analytic criteria, whereas experienced students prefer holistic criteria. As programme-focused assessment aims to encourage students to focus on long-term learning processes instead of short-term grade acquisition (Heeneman et al., 2015), task-specific criteria might be less suitable as these criteria tend to focus students on the task at hand, and less on what the student’s performance on this specific task tells about the student’s long-term development.

A programme perspective on transparency and meaning making of assessment criteria is helpful, because students should not be considered competent at judging the quality of their own and each other’s work from the start of the curriculum. Moreover, it is not realistic to expect students to become expert judges of their own work within the scope of a single module. But as the programme proceeds the students’ judgments of their own work should gradually reflect the (broad, holistic) programme learning outcomes and expectations of working professionals. The design of an assessment programme should give students insight in their learning and longitudinal development, ensure the main focus is on meaningful feedback to enable students to develop toward the programme learning outcomes (Bok et al., 2013a). If students

are to make meaning out of programme learning outcomes, then processes of feed up, feedback, and feedforward require a dialogue between students and teachers or between students and peers (Nicol and Macfarlane-Dick, 2006). Boud (2000) therefore argues that assessment should move away from the exclusive domain of the teacher/assessor into the hands of learners. Peer assessment could be implemented purposefully at different stages of the assessment programme. When they are still learning to become expert judges of quality, students need structure, and guidance when assessing their peers' work. Altogether, a programme perspective to transparency and meaning making of assessment criteria shows how students' meaning making process need to be purposefully guided over the period of the entire curriculum, and (formative) assessment moments, and evaluative judgment experiences should be purposefully planned and used.

## TRANSPARENCY, MEANING MAKING, AND SELF-REGULATORY SKILLS

Assessment and self-regulatory skills are intertwined in different ways. For instance, as Wiliam (2011) argues, an important aspect of formative assessment is activating learners as the owners of their learning process. In the same vein, Brown and Glover (2006) identified three levels of feedback: that which provides information about a performance; that which provides explanation of expected standards; and that which enables learners to self-regulate future performances. Also, Clark (2012) specifically links formative feedback to self-regulation, indicating that the objective of formative feedback is to support self-regulated learning and give the learner the power to steer one's own learning (p. 210). Furthermore, the use of specific assessment instruments may also have impact on students' self-regulation. As an example, Panadero and Romero (2014) examined the effects of using rubrics on students' self-regulation and concluded that "it is probable that the use of rubrics has a considerable impact on self-regulation, as its use promotes the strategies that have been shown to have the biggest effect on self-regulation interventions: planning, monitoring and evaluation" (p. 141). In other words, assessment and specific assessment instruments may foster students' self-regulation. Zimmerman (e.g., 2000) distinguishes three cyclical phases of self-regulation, that is, the forethought phase (occurs before efforts to learn), the performance phase (occurs during behavioral implementation), and the self-reflection phase (occurs after each learning effort). Especially in the first and third phase, assessment and assessment criteria may play a significant role. In the forethought phase, important processes are goal setting and outcome expectations. Even though very explicit and analytic assessment criteria may make clear to students what is expected of them, it may also produce students who are more dependent on their teachers and may weaken rather than strengthen the development of learner autonomy (Torrance, 2007). Autonomy may be understood as the ability to take care of one's own learning (Panadero and Broadbent, 2018). Consequently, transparency of assessment criteria may be detrimental for the development of students' self-regulatory skills because it may prevent them from

choosing their own learning goals and assembling their own learning path. In the self-reflection phase, self-evaluation (i.e., self-assessment) and causal attributions are main processes, and may be based on the same assessment criteria and standards. When we zoom in on meaning making of assessment criteria, meaningful assessment criteria will probably make it easier for students to formulate personal learning goals in the forethought phase of self-regulation as well as to self-evaluate after the learning efforts. Thus, we argue that meaningful assessment criteria may challenge students to regulate their own learning and increase their autonomy. Furthermore, a high level of self-regulatory skills may have impact on the way students deal with and interpret assessment criteria and standards. Panadero and Broadbent (2018, p. 82) argue that students who know how to self-regulate and to judge their own work can be autonomous and have more opportunities to develop evaluative judgement capacities (i.e., the ability to assess the quality of a piece of work).

So far, we addressed the relation between assessment (including transparency and meaning making) and self-regulatory skills mainly at the module level. But also at the programme level assessment and self-regulatory skills have a reciprocal relation. Zimmerman (2002) argues that self-regulation is important because a major function of education is the development of lifelong learning skills. Lifelong learning skills are necessary during an educational programme but also afterwards, in professional life. Assessment criteria and standards concerning the programme learning outcomes (as well as specific modules) should be meaningful in the sense that students should be able to grasp what these learning outcomes (i.e., assessment criteria and standards at the programme level) may mean for their own learning path and their professional development. Students' meaning making process of assessment criteria at the programme level may contribute to their ability to make choices and to become the professional they are aiming for. One of the goals of an entire curriculum and assessment programme can be to foster students' self-regulatory skills. An assessment programme hardly fosters these skills if the teachers tell students what to do and what to aim for. Instead, an assessment programme should reward students for identifying gaps in their abilities and developing effective ways to correct those gaps (Dannefer and Henson, 2007). In other words, an assessment programme that allows students to set their own learning goals and to formulate their personal assessment criteria and standards more explicitly calls for self-regulatory skills. Only a very small number of studies in a review on teachers' formative assessment practices (Gulikens and Baartman, 2017) showed examples of teachers allowing their students to set their own learning goals or allow students' learning goals to develop and change throughout a course or longer learning trajectory (e.g., Parr and Limbrick, 2010; Kearney, 2013; Lorente and Kirk, 2013; Hawe and Dixon, 2014). Concluding, we argue that an assessment programme should activate learners as the owner of their own learning process and allow learners to create meaningfulness of assessment criteria and standards by combining their own learning goals with available holistic assessment criteria and standards.



| Academic year 1   |   | Academic year 2  |   | Academic year 3   |  |
|---|---|--|---|---|--|
| Semester 1  | Semester 2  | Semester 3   | Semester 4  | Semester 5  | Semester 6                                 |
| <b>Professional Theory and Impact 1</b><br>Key task 1 t/m7<br>5 EC                    |   | <b>Professional Theory and Impact 2</b><br>Key task 1 t/m7<br>5 EC                     |   | <b>Professional Theory and Impact 3</b><br>Key task 1 t/m7<br>5 EC                    |  |
| <b>Teaching in vocational and professional education</b><br>Key task 1, 4, 5<br>10 EC | <b>Perspectives on vocational and professional education 1</b><br>Key task 1, 2, 3, 4, 5<br>10 EC | <b>Assessing in vocational and professional education</b><br>Key task 2, 4, 5<br>10 EC | <b>Designing vocational and professional education</b><br>Key task 3, 4, 5<br>10 EC | <b>Perspectives on vocational and professional education 2</b><br>Key task 5<br>10 EC | <b>Masterthesis</b><br>Key task 6<br>10 EC |
| <b>Practice-based research 1</b><br>keytask 6<br>5 EC                                 |   | <b>Practice-based research 2</b><br>keytask 6<br>5 EC                                  |   | <b>Practice-based research 3</b><br>keytask 6<br>5 EC                                 |  |

**FIGURE 1** | Master's Expert Teacher of Vocational Education Curriculum.

## METHODOLOGY

This study can be characterized as an in-depth single case study (Yin, 2014), which serves to empirically explore how designers, teachers, and students experience meaning making of assessment criteria and the role of self-regulation. In this single case study, we explored the design and implementation of a Master's programme in The Netherlands: the Master's Expert Teacher of Vocational Education (METVE). The METVE is a relatively new Master's programme and currently running for the third year (including a pilot year with 5 students). The design of this Master's programme (partly) includes the arguments about meaningfulness and self-regulation at the module level and programme level discussed above. Therefore, the METVE provided an interesting case to explore meaning making of assessment criteria in practice. The aim was to explore how the transparency of the programme learning outcomes was perceived and to reveal advantages and disadvantages of implementing transparency in terms of creating meaningfulness and fostering self-regulatory skills at the module level and at the programme level.

### Context of the Study (METVE)

The METVE is a part-time Master's programme for teachers of vocational subjects working in preparatory secondary vocational education (VMBO), senior secondary vocational education (MBO) and higher vocational education (HBO). In order to enhance the quality of vocational education in the Netherlands, the Educational Council of the Netherlands (2013) recommended increasing the standards of teachers: 25% of teachers of vocational subjects must have a Master's degree.

For higher vocational education, the aim is that by 2020, 100% of all teachers have at least a Master's degree. The METVE was developed to reach this goal and started in 2015 with a small pilot group of 5 students, continuing in 2016 and 2017 (15 and 20 students, respectively). All METVE students have (many) years of working experience as a teacher in vocational education, a bachelor degree (or equivalent) in their own occupational field (e.g., nursing, business, engineering) and a teaching certificate. Working in vocational education themselves, the METVE students have varied experiences when it comes to assessing their own students. In vocational education, students are generally assessed using a combination of knowledge tests, practical demonstrations, and assessments in the workplace. Competence-based standards are determined at the national level for the different occupations (for a more elaborate explanation about assessment in Dutch vocational education, see Baartman and Gulikers, 2017). The curriculum design of the METVE can be characterized as follows. The METVE curriculum works toward seven core tasks of vocational teachers, which are defined as the programme goals (or attainment levels): guiding students in vocational education, assessing students in vocational education, designing learning environments, connecting learning in-school, and outside school settings, connecting subject knowledge with the profession, doing practice-based research and professional development as a teacher. The METVE curriculum is divided into 5 or 10 ECTS modules in which students work on authentic assignments representing one or more of the seven core tasks. **Figure 1** gives an impression of the various METVE modules and core tasks.

For the entire METVE curriculum a general rubric has been developed based on the Dublin Descriptors, developed

as part of the Bologna Declaration and the Framework for Qualifications of the European Higher Education Area. The Dublin Descriptors provide descriptions of the different levels of higher education, developed to improve transparency and comparability of qualifications across Europe. For the METVE curriculum, the level descriptions of Bachelor (the entry level) and Master (the intended end level) were used. The Dublin Descriptors refer to the following five dimensions: knowledge and understanding, applying knowledge and understanding, making judgements, communication and learning skills. The general rubric serves to monitor and guide METVE students' long-term development process from bachelor-level toward master-level, across the different modules of the METVE. An English translation of the general rubric can be found in **Table 1**. For the different modules of the METVE, this rubric has been specified or contextualized into assessment criteria—again in a holistic way—to represent the assignment of that module, and what for example bachelor-plus performance looks like for that assignment. Within all modules, METVE students are assessed on the core tasks which are central stage in the assessment criteria. For example, in the first module, the METVE students develop a lesson plan for their own students in which they strive to connect learning across different sites inside, and outside the school. METVE students are assessed on three core tasks based on their lesson plan: guiding their vocational students, connecting learning inside and outside school and connecting subject knowledge with the profession. METVE students do not receive a score for their assignment (the lesson plan), but three separate scores for the core tasks. This way, the core tasks are in plain sight when assessing METVE students and students and teachers can monitor METVE students development on the core tasks throughout the curriculum (cf. van der Vleuten et al., 2012).

## Participants

Interviews were carried out with three developers/teachers, three teachers, six 1st-year students and four 2nd-year students. The three developers were involved in the design process of the METVE and the discussions about the underlying rationale of the master's programme. They were interviewed both in their role as developer and in their role as teacher. The three teachers got involved in the METVE in a later stage and had one to 2 years of experience in teaching in the METVE. Teachers and students participated voluntarily and signed a consent form for their participation.

## Interviews

In-depth (group) interviews were used to explore the self-reported experiences of METVE students and teachers with regard to the assessment criteria and the experienced meaning-making and self-regulatory activities. The individual interviews with the developers/teachers lasted 1 hour. One teacher was interviewed individually (30 min), and two teachers were interviewed together (60 min), based on possibilities in their teaching schedules. The students were interviewed in two group interviews, one for the 1st-year students and one for the 2nd-year students. Student interviews lasted 1 hour and were conducted using Adobe Connect (virtual classroom), a digital system the

students were familiar with in their webinars. All interviewed were audiotaped and transcribed verbatim. The interviews were carried out by the first and second author together. Because the first author is one of the developers/teachers of the METVE, the second author took the lead in the interviews to guarantee independence and stimulate the participants to freely express their opinion. Interview questions were asked about three topics: (1) transparency and meaningfulness of assessment criteria for the different modules / assignments of the MEB, (2) transparency and meaningfulness of the entire assessment programme, and (3) how self-regulation and ownership are addressed and experienced. Examples of questions asked to the teachers are: "how do you work with the assessment criteria in your module" and "how do you experience the connections between the modules in terms of students' long term development"? Examples of questions asked to students are: "how do you experience the freedom of choice when it comes to the assignments" and "what do you do to get an idea of what is expected of you in an assignment"? Besides the interviews, documents about the METVE were collected, such as policy documents, course guides and assessment forms. Some developers/teachers referred to these documents in their interviews and provided digital versions of the documents after the interview.

## Analyses

Thematic data analysis was carried out in three rounds by the two authors collaboratively. Template analysis was used, which consists of a succession of coding templates and hierarchically structured themes that are applied to the data (Brooks et al., 2015). After the interviews with the first and second developer/teacher, the 1st year and 2nd year students had been carried out, a first version of the template was developed by the two authors collaboratively, based on their experiences during the first interviews. The three research questions were used as an analytic framework: separate thematic codes were developed for each of the research questions. All fragments were coded using Excel: each fragment could be assigned a theme for either one, two, or all three research questions using pull down menus. This way, some fragments could be assigned to multiple themes when applicable (in practice, fragments never applied to all three research questions).

This first version of the template was used to analyze the interview with developer 1. Both authors coded the interview independently and their analyses were discussed in a meeting, resulting in adaptations to the themes (e.g., definitions were sharpened and themes were added that emerged from the data). The meeting resulted in a second version of the template. Also, the interviews with the other three teachers and developer 3 were carried out, in which the researchers asked follow up questions on themes that had not become clear in the first round of the analysis. Using the second version of the template, both authors independently coded the 1st and 2nd year student interview and (again) the interview with developer 1. The analyses were again discussed in a meeting, resulting in only minor changes in the themes (sharpening definitions so all fragments fitted the description of the themes). This resulted in the third and

**TABLE 1 |** General rubric of the METVE (translated from Dutch).

| Dublin descriptor                    | Bachelor  | Bachelor +   | Master   | Master + (excellent)   |
|--------------------------------------|---|--|--|--|
| Knowledge and understanding          | Student has knowledge and understanding based on specialized handbooks. Describes theories separately.  | Student has knowledge and understanding, partly based on scientific literature. Compares different theories.   | Student has deep knowledge and understanding based on scientific literature. Connects theories leading to new (practical) insights.  | Student has profound knowledge and understanding based on a large body of scientific literature. Connects theories leading to new (scientific) theories.   |
| Applying knowledge and understanding | Student applies knowledge in simple educational problems. Applies a theories to his/her own educational context.  | Student applies knowledge in complex educational problems. Chooses theories that fit the own educational context.  | Student applies knowledge in complex, new and/or interdisciplinary educational problems. Chooses and argues the best fitting theories for the own educational context, based on a comparison of scientific literature.   | Student applies knowledge in complex, new, interdisciplinary educational problems that go beyond the own educational context. Chooses and argues the best fitting combination of theories based on multiple scientific sources and the own experience.   |
| Making judgments                     | Student analyses educational problems in a partly systematic way. Critically investigates the own educational context. Identifies strengths and weaknesses based on single theories and practical experience. | Student analyses educational problems in a mostly systematic way. Critically investigates the own educational and broader social context. Identifies strengths and weaknesses and makes informed judgments based on theories and practical experience. | Student analyses educational problems in systematic way. Critically investigates the own educational and broader social context, in relation to scientific theories. Identifies strengths and weaknesses and makes informed judgments based on (in)possibilities in practice and research. | Student makes a thorough and critical analysis of educational problems using multiple perspectives. Critically investigates the own educational and broader social context, comparing practical (im)possibilities and scientific theories. Choices are made based on scientific theories and validated in practice.                      |
| Communication                        | Student participates in the own educational institute. Communicates ideas and possible solutions with colleagues.   | Student participates in a network in the own educational institute and surrounding industry. Discusses ideas and possible solutions with colleagues and representatives from industry partners.  | Student participates in a network in the broad educational and industrial context. Discusses ideas and possible solutions, including underlying motives and considerations with colleagues, and representatives from industry partners.  | Student participates in a network in the broad educational and industrial context. Takes initiatives and shares knowledge, for example in publications, presentations and workshops. Discusses ideas and possible solutions, including underlying motives and considerations with colleagues and representatives from industry partners. |
| Learning skills                      | Student shows continuous development by means of reflection on his/her own professional performance in the broad context of vocational education.   | Student shows continuous development by comparing his/her own professional performance to theory (doing and thinking). Analyses the discrepancies between doing and thinking.  | Student shows continuous development with a high degree of autonomy. Learns by means of conceptualizing and concretizing.  | Student shows continuous development with a high degree of autonomy. Learns by means of conceptualizing and concretizing. Takes the initiative to learn from complex situations.   |

final version of the template, which was used by both authors independently to (re)analyze all interviews (see **Table 2** for the final template). Finally, both authors independently selected all fragments belonging to a theme (using the pull-down selection menu), re-read the fragments and made a summary of the theme together with some illustrating examples. This was done for all themes separately. In a meeting, the summaries of the themes belonging to research question 1 were discussed (4 themes). The authors read each other's summaries and made notes when they noticed (big) differences. The main question that guided the discussion was: do we see any results/conclusions that do not logically emerge from the data? Some differences between the authors did appear, mainly because of overlap between some of the themes. For the summaries for research questions 2 and 3 both authors again audited each other's summaries and made notes of differences they encountered. The first author used the discussion and the notes to make a final summary per theme, which was checked by the second author.

## RESULTS

The results are presented per research question and fragments are added as illustrations of the themes that appeared from the data.

### Fostering Meaning Making of Assessment Criteria at the Module Level

For the first research question “How can students’ meaning making at the module level be fostered,” four themes were derived from the data, related to (1) holistic assessment criteria, (2) meaning making by means of practical relevance, (3) meaning

making by evaluative experiences, and (4) making connections between the general rubric and assignments at the module level.

The interviews showed that in the design of the METVE programme, a holistic approach to assessment was deliberately chosen because of the diversity of the student population, who all work in different domains and levels of vocational education. Also, one developer explained how holistic criteria do more justice to the complexity of tasks students encounter in practice: “an educational argument is that we think if you take a more holistic view, well that is actually how the core tasks, how complex they are. So there is no recipe for carrying out vocational education. That recipe does not exist. So we cannot work out the assessment criteria from A to Z. Because they just not exist. So you have to assess holistically” (developer/teacher nr1).

Though the teachers seem to value the holistic assessment criteria, METVE students reacted in a more diverse way: “well, it may depend on me as a person ... I ... I think this broader framework and the fact we are not pushed into a certain direction, it also gives you the possibility to work out an assignment in your own way” (2nd year student nr1) or “I notice, but as a person I work in the technical domain ... and there you are pragmatic, I like to have a guiding principle to deliver something” (2nd year student nr2). METVE students sometimes seem to feel uncertain about what is expected of them. As one of the students described it: “You have freedom in how to do the assignments, but this freedom can also make you insecure, because you can’t exactly pin down what the purpose is of what you have to show” (2nd year student nr4). Both teachers and students still need to build up impressions of what good work might look like in all its possible varieties. One developer explained: “But students also have to build up these images, but we as teachers also need to build up the images, I experience” (developer/teacher nr2). This student agrees with the holistic assessment criteria, but also expresses his need for certainty: “But I think, when you use a holistic assessment model—I understand the goal of that very well—that you also need a further explanation of the module, or the goals of the module, from the start. So if you ask me, it is connected: you either get your information from the assessment criteria or it needs to be made clear that the assessment criteria do not contain all specific information” (2nd year student nr 3).

The diverse METVE student population and holistic assessment criteria bring us to the second theme, namely the increase of meaningfulness by contextualizing assignments to the METVE students’ work context. It is this connection to their own work field that makes the assignments relevant or meaningful, and indirectly, also increases the meaningfulness of the holistic assessment criteria. The module assignments explicitly require METVE students to explore developments in their own work context, for example by talking to colleagues, managers, and experts. METVE students thus address a relevant problem or question experienced in their own work context and they try to realize impact on their work context, for example by the products they develop: “we could make our own choice when it comes to the content of the assignment. I chose self-management, because I work with elderly clients. Yes, it was very meaningful for me.

**TABLE 2** | Final template used for data analysis.

|                         |  |
|-------------------------|--|
| Module level            | <p>Student and teacher experiences of holistic assessment criteria</p> <p>Meaning making by means of practical relevance</p> <p>Meaning making by organizing evaluative experiences</p> <p>Making connections between the general METVE rubric and assignments at the module level</p>                                   |
| Programme level         | <p>Design of the METVE curriculum</p> <p>Teacher and students activities to support meaning making of assessment criteria at the programme level</p> <p>Students get an impression of their development throughout the programme</p> <p>Conditions or prerequisites to achieve meaning making at the programme level</p> |
| Self-regulated learning | <p>Students’ own development at the module level</p> <p>Students’ own development at the programme level</p> <p>Experiences of lack of time</p> <p>Students’ professional role after graduation</p> <p>Making self-regulation more explicit</p> <p>Teachers’ support of self-regulation</p>                              |



I really experienced an added value ... I could also ask better questions to my pupils" (1st year student nr1). And another student: "You explore what is going on in your department, what relevant issues are, what you like to know more about, in collaboration with your team. So I discuss with my team which research question I address. And also in the module about assessment, I discussed what issues there are, what we like to have an answer for" (2nd year student nr2). This meaning making process goes two ways, from the METVE assignment to the work context of the METVE student and the other way around. Students thus move within the boundaries set by the assessment criteria and within these boundaries experience freedom to contextualize the assignments: "our choice is the topics ... I think, with all assignments, within the boundaries you are free to make choices" (2nd year student nr1). A METVE developer explained how they safeguard the boundaries the students move within, so choices METVE students make fit within the holistic assessment criteria: "we do tell them to go to their own school, talk to colleagues, managers, what is interesting. Sometimes they get a bit stuck, like my manager wants this, but I don't know if it is relevant to the METVE. So that is a step we safeguard, is it a relevant assignment" (developer/teacher nr2).

Themes 3 and 4 at the module level are related to each other and portray how teachers and students work on meaning making by means of evaluative experiences (theme 3). The goal of these evaluative experiences is to make connections between the assignments the students are working on, the assessment criteria and the general rubric (theme 4). Some evaluative experiences were explicitly designed in the curriculum, for example the formative moments during the modules in order to give feedback while students are still in the midst of the meaning making process and can still make choices and adaptation in their assignments. One of the developers explained: "well, for all modules ... it is a formative process. Student work from moment zero toward the end result that will be assessed. So it is not just some separate small assignments, that you first get assignment 1, and then 3, and then 6 and all small assignments together result in ... no, they actually work all the time toward that end product. So all feedback they get is about the end product" (developer/teacher nr1). Or: "it is an holistic assessment about the three [formative part-assignments], in which I explicitly give the message to students that the part assignments have a formative goal, and that I give feedback to part assignment 1 based on the assessment criteria and the Dublin Descriptors so they can show growth within my module" (developer/teacher nr3). Other evaluative experiences were not explicitly designed and depend on individual METVE teachers. Examples of evaluative experiences used in the lessons are: discussions of student assignments using the assessment criteria, peer feedback activities, peer group intervision, teacher feedback, and modeling how you assess as a teacher. These quotes show how METVE teachers and students tell about the evaluative experiences:

"I think it is very worthwhile, to do it often. And it does happen in some instances. In module XXX the teacher projected a student's piece of work on the smart board, and well, how you would assess

it based on the assessment criteria. We could do that more often, it would help immensely" (1st year student nr4)

"Some time ago we discussed the rubrics during the XXX lessons, because we are going to do peer assessments ... and actually we worked out the rubric in pairs. That was very clarifying because you actually, because per theme [assessment criterion] you discuss well, how you would assess someone as a peer assessor. I found that a very interesting addition, to make the holistic more concrete" (2nd year student nr2).

"I used some activities in the lessons, I let them compare some examples. They had to bring their own assignment and in small groups, using the assessment criteria—not really the rubric—they looked for good examples of what assessment criteria might look like. And they made big posters of the examples" (developer/teacher nr3).

METVE students express the value of the evaluative experiences, but they would like to have more evaluative experiences during the lessons, and especially at the beginning of a module to get an impression of what is expected: "I would appreciate it very much if it were at the beginning of the module. And I would like to discuss it in class. Like look, this is the level you are at now, and this is what you are heading for, the ultimate goal is master level" (1st year student nr2). There is a demand to explicitly discuss the assessment criteria, because the general rubric contains concepts that students find hard to grasp: "well, as a student you apparently like to be taken by the hand a bit, so in the lessons you like the teachers to help you, in the right direction ... and then you assume it is all right. And the assessment criteria I think, I see a number of sentences, but what is exactly meant by them?" (2nd year student nr4). The students also expressed the limits of peer feedback: "it is also about that you have to know how to interpret the assessment form. Because when you have not made anything yet, and if you do not know how a teacher would assess it, then it is difficult to grasp. Because we also gave each other feedback and one of us used the assessment form and the other three did not. You have to learn to read it" (1st year student nr6).

The goal of the evaluative experiences, but also the design of the METVE programme, is to make connections between student assignments, the assessment criteria and the general rubric. These connections were meant to increase meaningfulness to METVE students: they can analyze how their specific assignment, choices, and work context relate to the assessment criteria and thus whether they comply with the assessment criteria and the required (master) level. In practice, METVE students do use the assessment criteria to find out what is expected of them at the module level, for example by reading the module guide, and assessment criteria and comparing them to their own work. Students also expressed they rely on the teacher: "I think I do not use it [the assessment form] very often, less than I should ... I think I just use the lessons to know what is meant by the assignments, and then I just get to work. And actually, I just use it at the end as a kind of checklist to check whether I did everything that is expected. But during the lessons you get so much feedback and input from the teachers, I actually lean more from that than from the assessment form itself" (2nd year



student nr4). In the design of the METVE, the general rubric has been translated in assessment criteria for the different modules. These assessment criteria are more specific and concrete and are thus more (directly) meaningful to students. For students, the relationship between their assignments and the general rubric is not always clear. One teacher told: “well, they do read the rubric ... or I think so ... but in the end they look more at the part assignment and the criteria for the assignments. Even if we made an explicit connection between the part assignment and how they are linked to the rubric. But they do not really look at that ... well, it is more contextualized ... it is less general” (teacher nr3).

## Fostering Meaning Making at the Programme Level

Four themes illustrate how meaning making of assessment criteria is fostered at the programme level: (1) by the design of the METVE programme, (2) by teacher and student activities, (3) because students get an impression of their development throughout the programme, and (4) conditions to be met if meaning making is to take place at the programme level.

First, meaning making at the programme level is fostered by the design of the METVE programme. A programmatic approach to assessment is used (van der Vleuten et al., 2012) within constraints such as the demand for a modular curriculum: “Or course, we had the idea of gradual development in the curriculum, a development line. That was the first dilemma in our curriculum, because we wanted a nice progression and an increase in complexity, while actually the demand was that students should be able to do separate modules. That you do only one module. So that is kind of tension in our curriculum” (developer/teacher nr1). Important elements of the METVE design that foster students’ meaning making and long-term learning processes are the general rubric to assess student work in the various modules and the fact that students can show growth on the core tasks throughout the curriculum: “so that is the thread of our curriculum structure. And the core tasks come back several times, you can develop toward master level. Core tasks 1/2/3 are addressed very prominently only one time, in their own module, so you have to show the master level immediately. But core tasks 4 and 5, and 6 in practical research, they come back three times at least ... so you can grow” (developer/teacher nr2). And: “Well, we made a rubric for the entire master ... and we said, we always assess the core tasks, in all modules. So the students develop a product and you could say, we assess the product, but we assess the three core tasks that are addressed in that product” (developer/teacher nr2). The METVE students recognize the design of the programme, but also add that even more connections could be made between the modules, to make the programme even more meaningful to them: “but you could also stress the connections between the modules ... because practical research, it would be good if you use the topic of the module about guidance, that you do research on that topic. So I would stress the connections much more, that you can use practical research in the other modules. I think, now we have three separate modules whereas there are so many connections. Now I see that, I think ... well, you

should also stress it at the start of the METVE (1st year student nr2).

The second theme illustrates how the programme design alone does not guarantee meaningfulness of assessment criteria. It needs to be purposefully designed *and* realized by the teachers. METVE students—especially the 1st year students—tell they find it difficult to look far ahead: “well, let’s be honest, we are starting students, really, we are not trying to find out what you have to do three years from now” (1st year student nr2). This teacher also realizes 1st year students cannot have an overview of the entire programme: “this is how I tell students in my module [end 2nd year], because I think at that moment they can understand. Because I think it is quite complex if you tell this at moment zero. Because they do not understand the programmatic perspective yet” (developer/teacher nr1). This theme thus also seems to show differences between the needs of 1st year and 2nd year students, which has implication for programme design and teacher activities (e.g., a full programmatic perspective might be too much to ask from 1st year students).

METVE teachers mentioned some strategies they use to foster students’ meaning making at the programme level. Teachers not only give feedback on current assignments, but also give feed forward that indicates what students have to do to improve toward the master’s level (as described in the overall rubric): “what I do, when I assess bachelor-plus level, then I write down what they have to do to reach the master’s level. So even if only bachelor-plus is required, I always add this, for the master level” (developer/teacher nr2). Also, teachers try to stimulate students to make connections between the different modules, in which they work on their core tasks and grow toward the master level: “in the entire METVE you want to guide them toward a certain level. And if you cannot see what your module contributes to this development ... if you cannot put that next to the contribution of the other modules ... well, that is not handy” (teacher nr1). And: “what we do is, we ask students at the start of a module, well bring the assessment form of the previous module. They get a lot of feedback, and we ask students what are your strong and weak points if you look at your last assignment and feedback. And what does that mean, what are you going to work on now” (developer/teacher nr3).

To work on meaning making of assessment criteria at the programme level, students need to have an impression of their own development toward the master level (theme 3). Only then, students can connect their own development to the assignments of the modules and their choices of what to work on and improve. METVE students and teachers described how they notice development and growth. For example: “You notice that they strive for quality, you see assumptions and arguments ... that they suddenly realize, why am I doing this? When I see that...” (developer/teacher nr1). And: “actually, they are too successful, that is a criterion for me ... [...] ... they get more tasks. So you notice ... at their workplace. They are taken seriously as a discussion partner, they notice more ... I am not sure whether they are more interfering with matters ... you see their workload increases” (developer/teacher nr1). The following examples show how METVE students notice their own development (or not):

“Well yes, if you have reached the required level ... or if it is only for this assignment, that is the question of course. So overall ... actually I do not have a clear impression for myself” (1st year student nr4).

“For me ... that I learn to use tacit knowledge, so I get a grip on the domain, on being a vocational teacher. Just because you notice much better what you are doing ... I notice that my work really develops, I make more deliberate choices” (2nd year student nr4).

“Well ... I maybe find the results of the assignment less valuable. What I am looking for is my performance at the workplace. I think I should act at another level in my organization. And this has nothing to do with whether I finish the assignments and get a pass. So that part, the credit points you get ... that is nice and I know I have to do it. But they don't mean too much for where I am standing now. (2nd year student nr1).

Finally, the last theme shows a number of prerequisites for meaning making of assessment criteria at the programme level. The METVE programme has ran for just two years and the teachers still need to develop a certain routine: “Now we have to develop a routine, as a team. And you have to learn to carry out the design, you have to be on the same wavelength. That is phase we are in now” (developer/teacher nr1). Developers, teachers and students agreed that they are still searching for the different possible interpretations of assessment criteria: what choices can be made, what are the different variations of good work. Teachers develop these impressions during their first (and subsequent) years of teaching, in which they encounter many variations of student work. This also raises some issues with regard to new teachers who start working as a teacher in the METVE programme, because meaning making at the programme level requires a full overview of the curriculum. In the interviews, teachers told they are better able to guide students and help them make meaning of the criteria after their first year of teaching: “well, I notice, now we do it the second time, that I can be sharper in dialogues with students ... I am better able to guide the discussions because I formed a picture of what they can choose ... you can give examples” (developer/teacher nr2). Also, to really work toward the programme goals in all modules requires that teachers have an overview of the entire curriculum, share these goals and know what is going on in other modules. In other words, as one of the developers said: “so it is a team effort and not an effort of teacher who all do their own little part” (developer/teacher nr3).

## The Contribution of Self-Regulation to Students' Meaning Making Process

For the third research question, “How can self-regulation contribute to students' meaning making process?” six themes were derived from the data, related to (1) students' own development at the module level, (2) students' own development at the programme level, (3) lack of time, (4) students' professional role after graduation, (5) making self-regulation more explicit, and (6) supporting self-regulation.

The first theme, students' own development at the module level, refers to the choices students make for their assignments. These choices are not always based on what their professional context is asking, but also on how they want to develop themselves and what they want to learn. So, at the module level, students are challenged to show ownership. As one of the teachers explained this: “We do this by asking them to find domain experts [for a specific assignment]... and what you encounter is the quality of the expert (...) that way we try ownership...” (developer/teacher nr2). Students pick up that responsibility and are sometimes proactive: “Well, I complete an assignment, hand it in and then I ask feedback. Sometimes I ask feedback from my critical friends [peers], which is often supporting, some advice together. Or feedback from the teacher and based on that feedback I get a notion of what is actually meant.” (2nd year student nr4).

The METVE teachers acknowledge how important it is that students pay attention to their development at the programme level, which is the second theme concerning self-regulation. Developer/teacher nr2 stated: “When we designed the programme, we said that in particular at Master's level, and in particular where it concerns a teachers' programme, it is very important that students are able to self-assess, and that they are able to handle such an assessment instrument [the general rubric].” Not all students are able to show self-regulation at the programme level. There are big individual differences. Teacher 2 explained: “I think that the majority [of the students], not everyone, are focused on their development at Master's level. And how they can contribute to their professional practice.” Developer/teacher nr1 also noticed that some students show quite passive behavior throughout the programme: “Sometimes we [teacher and student] joke about it, like hey, you behave like a student again, what's this? We discuss that with them, like well, you can act as a sort of passive student and the teachers says this and that,... so that is sometimes a sort of discussion, and we say that is not the way it works...” As has been stated before, the freedom students have for completing assignments the way they want can make them a bit insecure and, consequently, quite passive. Several students thus indicated that they want more explicit attention for self-regulation at the programme level. For example, 2nd year student nr2 indicated: “I think I know my strong and weak points, but we miss some sort of career counselor.” So, the programme allows for self-regulation at the programme level, but it is not explicitly addressed and supported. It is up to the individual student whether they will accept the challenge.

The third theme, lack of time, is related to what was just described. Students indicate that at the beginning of the year they were asked to write down what they wanted to accomplish, but because the programme is perceived as very time consuming and difficult to combine with a job and family, they revealed that they did not have the time to reflect on these personal learning goals. Some students are happy if they can just do what the teachers tell them and they feel that setting their own learning goals costs extra time they don't have: “Studying 20 h a week in combination with a job, that is just hard to formulate your own learning goals, etcetera. So, indeed, it is like okay, I do that module, tak

tak, preparing the face-to-face meeting for next week. It is very tight schedule, check stuff, prepare your own work, and then it is already Monday again.” (1st year student nr5).

The fourth theme, students’ professional role after graduation, was only mentioned in a few occasions. Basically, the message was that students’ professional role after graduation is hardly addressed: “[students] are really looking for their role, like how can I get this all together, and what do other people gain from me as an expert teacher at Master’s level? (...) They find that really hard, because their role is new, no-one really knows yet, so the tasks you [the expert teacher after graduation] gets, do they fit?” (developer/teacher nr2). Students and teachers recognize this: “We actually never discussed each others’ personal learning goals and never reflected on how they fit in the METVE, and how come that you have these goals, and does it have something to do with the opportunities you have at work... and then I come back to what am I going to do after I finished the METVE, for me in my higher education institute. My personal goals are really related to that” (2nd year student nr3) and “Yes, because they [students] have no idea which role they will get in the future, what they are able to do and know when they are Master’s teacher in vocational education. But we could ask them to create that image, based on what they learn now, what they think, what they get out of the programme, what kind of role they see for themselves, when they are that Master’s teacher. That will put it more in perspective, yes, maybe learn from it and a career or something, an orientation on the career (...) and getting the self-regulation from that” (teacher nr2).

The last two themes, make self-regulation more explicit and support of students’ self-regulation, are related in the sense that the teachers have an important role in this. In the METVE, there is some attention for the development of self-regulation in the design of the programme, but in reality, this is rather implicit. Teachers find it important to explicitly communicate this expectation, that is, that METVE students take up self-regulation at the programme and module level. “Yeah, I had this conversation [with a student] (...) and I heard her saying: what do I have to do? And I said, I think we offer boundaries in which learning takes place, so what do you want to learn? That is nice of METVE, students can fill that in for themselves. But I think we stated that too implicitly” (teacher nr1), and “I think that for self-regulation, that starts with expectations... what do I expect of students at the end of module XXX, what are the attainment levels, and are students allowed to show a more or lesser degree of self-regulation...” (teacher nr2), and “We have this vision, but we don’t show that explicitly, and that makes that students not always pick that [self-regulation] up, and we just say, yeah, they are Master students so...” (teacher nr2).

So, in the METVE, the development of self-regulation could have been more supported. Students are not able to develop their self-regulation at the programme level automatically. Implicitly, teachers expect students to be able to when they enroll and acknowledge that this can be improved. “Maybe we should guide that in a better way or build that up, because now we kind of let that go in my opinion, and assume that they will do that, and maybe we could say that first we take them by the hand and then... give them more freedom” (teacher nr1) and “Well, they

[students] can ask feedback on the things they are working on, and they do that, but maybe they have to ask specific feedback questions. Because in the design we said that before you ask feedback from your teacher, you have to formulate a specific feedback question. Don’t just go to the teacher and say, here is my assignment, what do think of it? (...) They find that hard. It demands a bit of self-evaluative judgment, because you have to be able to estimate what your strong and weak points are to be able to ask a good feedback question.” (developer/teacher nr3) Teachers do not always feel that they are able to support students’ self-regulation at the programme level sufficiently. On the one hand, they need time, a good overview of the curriculum, and a notion of what students will do after they are graduated. On the other hand, because the programme is still quite new, at the moment they are mainly busy with improving the quality of the content of their own module. As teacher nr3 puts it: “I always think, in the future, when I have more hold on it, next year, I can improve. Now I am less focused on... stimulating students to formulate their personal goals... I didn’t even think of that... but if I would have wanted to do that... I think let’s first just manage my own content before I go outside of that. I just can’t handle that at the moment.” (teacher nr3).

## CONCLUSION AND DISCUSSION

The research questions that guided this exploratory case study focused on fostering meaning making of assessment criteria at the module level and the programme level, and the role of self-regulation in this meaning making process. We presented a single exploratory case study in order to explore processes of meaning making of assessment criteria by curriculum designers, teachers, and students. It needs to be noted that this study was conducted in a teacher education context and the results might not apply to other higher education courses. Also, we used interviews in which the participants self-reported about their experiences with regard to the meaning making of assessment criteria, which might have affected the results.

Our study explored how meaning making takes place at the module level (research question 1) by using holistic assessment criteria which allow students to make choices within the boundaries set by the assessment criteria. Comparable to Sadler’s (1989) argument, the METVE teachers seem to value holistic assessment criteria, but the counter side may be that students seem to experience insecurity as holistic criteria provide less guidance on what is expected. In this respect, previous research on the use of rubrics (e.g., Jonsson and Svingby, 2007) shows that value of task specific rubrics. When holistic criteria are used, meaning making at the module level should thus be fostered by creating evaluative experiences, such as comparing examples, peer feedback and modeling practices by the teacher. The evaluative experiences mentioned by the METVE teachers show they go beyond telling and showing desired learning outcomes, as is recommended by several authors (cf. Fluckiger et al., 2010; Willis, 2011; Hawe and Dixon, 2014).

This study also indicates that meaning making seems to be more difficult at the programme level (research question 2).

The design of the METVE programme aims to foster meaning making at the programme level by using a general holistic rubric that is used in all modules, and by assessing the core tasks throughout the curriculum, so students can show growth (van der Vleuten et al., 2012; cf. Bok et al., 2013a). However, the design of a curriculum alone cannot ensure meaning making at the programme level. As research on evaluative judgment shows, fostering students' capacity for evaluative judgment requires pedagogic practices, in this case focused on the programme level and students' long term learning process toward the programme goals. The METVE teachers give feedback and feed forward using the general rubric and stimulate students to take feedback from one module to another. This approach to giving feedback is also advocated by Hughes et al. (2015) to stimulate student learning beyond the module level. Other teacher and student activities seem to center around creating evaluative experiences for students (Boud et al., 2018). Examples found in this study are peer assessment, dialogues and intervention activities. More research is needed, however, to explore to what extent these activities really focus students' attention on the programme goals or graduate learning outcomes. Students seem to tend to focus on the upcoming assignment and less on their development throughout the programme and beyond.

When it comes to self-regulation (research question 3), we have seen that the METVE programme (including the assessment criteria and standards) is designed in such a way that the design allows for self-regulation at the programme level. If students want to formulate and evaluate their own learning goals, they can (within the boundaries of the holistic assessment criteria). However, it is up to the individual student whether they will take on this challenge and students seem not to do this spontaneously (only guided by a teacher). Students also perceive it as something extra for which they actually don't have enough time and the freedom they have in the programme sometimes makes them feel insecure. Furthermore, self-regulation at the programme level is not explicitly supported in the METVE, for example by a study coach. That makes it more difficult for some students to create meaningfulness of assessment criteria and standards and use this meaning making process to regulate their own learning process toward the role they want to fulfill after completing the Master's programme.

In general, this study provides some practical implications for the design of higher education courses and (starting) teacher professional development in higher education when it comes to fostering meaning making of assessment criteria at the module level and the programme level. First, our case study seems to indicate some prerequisites for meaning making to happen, especially at the programme level. In order to design and carry out evaluative activities that foster meaning making at the programme level, teachers—just as students—need to develop an overview of all possible varieties of student work that fit within the holistic assessment criteria. Also, teachers need to be familiar with the entire curriculum—and not just their “own” modules—to be able to give feedback and feed forward across modules. Teaching thus becomes a team effort instead of an individual activity (cf. Jessop and

Tomas, 2017). Just as students can discuss different examples of student work to develop a more diverse picture of what quality might entail in diverse vocational situations, (starting) teachers could do the same in professional development activities and discussions when judging student work (Sadler, 1989; Boud et al., 2018).

Second, curriculum designers (in higher education) could take into account the programme perspective from the start of their design process. They could design a sequence of assessment methods that increasingly stimulate students' capacity for evaluative judgment. They could design evaluative experiences at the programme level and even beyond, by addressing the role of the students after graduation (Boud, 2000), when professionals also have to be able to judge what good work entails. Evaluative judgment concerns the evolving ability to engage with quality criteria and make informed judgments about one's own work and that of others (Boud, 2000; Sadler, 2009; Carless, 2015; Panadero and Broadbent, 2018). During their engagement in the curriculum with different assignments and activities organized by the teachers, students can gradually develop a sense of quality, like in Carless' study (2015) by presentations and peer feedback resulting in an increase in transparency by exemplifying how criteria and standards can be applied in diverse products.

Third, in order to foster meaning making of assessment criteria at the programme level, students can be stimulated to take a more active role in meaning making processes. This study revealed that students—in this case busy working students—tend to work with the assessment criteria on their own (for example by using the assessment criteria as a checklist). We believe that for students (as well as teachers), meaning making could benefit from collaborative processes like intervention, peer feedback, and dialogues (e.g., Sadler, 1989; Carless, 2015). This also implies that, despite time pressure and insecurity, higher education students should develop an attitude to deal with insecurity and work on meaning making and self-regulation in collaboration. Although it seems quite hard for students, we argue that this is necessary for students to become lifelong learners.

Finally, a discussion about analytic vs. holistic assessment criteria seems warranted. In our exploratory case study, holistic criteria were used to foster meaning making and clarity of student progress at the programme level. Holistic criteria leave room for meaning making and self-regulation (Sadler, 2009). On the other hand, assessment instruments with a more analytical perspective may be beneficial for students as well (e.g., Weigle, 2002; Jonsson and Svingby, 2007), because task specific rubrics and analytic criteria may provide more specific diagnostic information for improvement that can be used by teachers and students. Govaerts et al. (2005) found that students' experiences with regard to more analytic vs. holistic assessment criteria varied depending on their experience. Beginning students prefer analytic assessment criteria because they provide clear guidance for learning, whereas more experienced students prefer holistic criteria as they perceive analytic criteria to be checklists that do not really capture what is important in professional practice. Our study seems to indicate a similar distinction. This exploratory case study thus seems to indicate that what is needed with regard to assessment criteria might



be different at the beginning of a curriculum than at the end. Again, this seems to advocate a programme perspective on curriculum design, with more analytic assessment criteria and specific meaning making activities at the beginning, and more holistic criteria, options of task customization and peer feedback toward the end of the curriculum. Future research could take up the challenge to explore how students experience their engagement with assessment criteria throughout the curriculum. Also, because the current study was a single exploratory case study, more research is needed to further investigate the programme perspective on assessment criteria, for example on evaluative judgment and how students' capacity for evaluative judgment can be fostered throughout the curriculum, and how students' attention can be geared toward the programme goals and the future profession instead of the short-term upcoming assignment.

## REFERENCES

- Baartman, L. K. J., and Gulikers, J. T. M. (2017). "Assessment in Dutch vocational education: overview and tensions of the past 15 years," in *Enhancing Teaching and Learning in the Dutch Vocational Education System: Reforms Enacted*, eds E. de Bruijn, S. Billett, and J. Onstenk (Springer), 245–266.
- Black, P., and William, D. (1998). Assessment and classroom learning. *Assess. Educ.* 5, 7–74. doi: 10.1080/0969595980050102
- Bloxham, S., and Campbell, L. (2010). Generating dialogue in assessment feedback: exploring the use of interactive cover sheets. *Assess. Eval. Higher Educ.* 35, 291–300. doi: 10.1080/02602931003650045
- Bok, H. G., Teunissen, P. W., Favier, R. P., Rietbroek, N. J., Theyse, L. F., Brommer, H., et al. (2013a). Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Med. Educ.* 13:123. doi: 10.1186/1472-6920-13-123
- Bok, H. G., Teunissen, P. W., Spruijt, A., Fokkema, J. P. I., van Beukelen, P., Jaarsma, D. A., et al. (2013b). Clarifying students' feedback-seeking behaviour in clinical clerkships. *Med. Educ.* 47, 282–291. doi: 10.1111/medu.12054
- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Stud. Cont. Educ.* 22, 151–167. doi: 10.1080/713695728
- Boud, D., Ajjawi, R., Dawson, P., and Tai, J. (2018). *Developing Evaluative Judgement in Higher Education. Assessment for Knowing and Producing Quality Work*. London: Routledge. doi: 10.4324/9781315109251
- Boud, D., and Falchikov, N. (2006). Aligning assessment with long-term learning. *Assess. Eval. Higher Educ.* 31, 399–413. doi: 10.1080/02602930600679050
- Brooks, J., McCluskey, S., Turley, E., and King, N. (2015). The utility of template analysis in qualitative psychology research. *Qual. Res. Psychol.* 12, 202–222. doi: 10.1080/14780887.2014.955224
- Brown, E., and Glover, C. (2006). "Evaluating written feedback," in *Innovative Assessment in Higher Education*, eds C. Bryan and K. Clegg (London: Routledge), 81–91.
- Carless, D. (2015). Exploring learning-oriented assessment processes. *High. Educ.* 69, 963–976. doi: 10.1007/s10734-014-9816-z
- Clark, I. (2012). Formative assessment: assessment is for self-regulated learning. *Educ. Psychol. Rev.* 24, 205–249. doi: 10.1007/s10648-011-9191-6
- Conway, R. (2011). *Owning Their Learning: Using assessment for Learning to Help Students Assume Responsibility for Planning, (some) Teaching and Evaluation*. Teaching History, 51.
- Dannefer, E. F., and Henson, L. C. (2007). The portfolio approach to competency-based assessment at the cleveland clinic lerner college of medicine. *Acad. Med.* 82, 493–502. doi: 10.1097/ACM.0b013e31803ead30
- Fluckiger, J., Vigil, Y. T. Y., Pasco, R., and Danielson, K. (2010). Formative feedback: Involving students as partners in assessment to enhance learning. *Coll. Teach.* 58, 136–140. doi: 10.1080/87567555.2010.484031
- Govaerts, M. J. B., Van der Vleuten, C. P. M., Schuwirth, L. W. T., and Muijtjens, A. M. M. (2005). The use of observational diaries in in-training evaluation: student perceptions. *Adv. Health Sci. Educ.* 10, 171–188. doi: 10.1007/s10459-005-0398-5
- Gulikers, J., and Baartman, L. K. J. (2017). *Doelgericht Professionaliseren: Formatieve Toetspraktijken Met Effect! Wat DOET de Docent in de Klas? [Goal Directed Professionalisation: Formative Assessment Practices With Effect! What Do Teachers Do in the Classroom]*. Netherlands Initiative for Education Research. Report number: NRO PPO 405-15-722.
- Hartley, P., and Whitfield, R. (2011). The case for programme-focused assessment. *Educ. Dev.* 12, 8–12.
- Hawe, E. M., and Dixon, H. R. (2014). Building students' evaluative and productive expertise in the writing classroom. *Assess. Writ.* 19, 66–79. doi: 10.1016/j.asw.2013.11.004
- Heeneman, S., Oudkerk Pool, A., Schuwirth, L. W., Vleuten, C. P., and Driessen, E. W. (2015). The impact of programmatic assessment on student learning: theory versus practice. *Med. Educ.* 49, 487–498. doi: 10.1111/medu.12645
- Hughes, G., Smith, H., and Creese, B. (2015). Not seeing the wood for the trees: developing a feedback analysis tool to explore feed forward in modularised programmes. *Assess. Eval. High. Educ.* 40, 1079–1094. doi: 10.1080/02602938.2014.969193
- Jessop, T., and Tomas, C. (2017). The implications of programme assessment patterns for student learning. *Assess. Eval. High. Educ.* 42, 990–999. doi: 10.1080/02602938.2016.1217501
- 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
- Kearney, S. (2013). Improving engagement: the use of 'authentic self-and peer-assessment for learning' to enhance the student learning experience. *Assess. Eval. High. Educ.* 38, 875–891. doi: 10.1080/02602938.2012.751963
- Lokhoff, J., Wegewijs, B., Durkin, K., Wagenaar, R., Gonzales, J., Isaacs, A. K., et al. (2010). *A Tuning Guide to Formulating Degree Programme Profiles. Including Programmes Competences and Programme Learning Outcomes*. Bilbao; Groningen; Hague: Nuffic/TUNING Association.
- Lorente, E., and Kirk, D. (2013). Alternative democratic assessment in PETE: an action-research study exploring risks, challenges and solutions. *Sport Educ. Soc.* 18, 77–96. doi: 10.1080/13573322.2012.713860
- Nicol, D. J., and Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Stud. High. Educ.* 31, 199–218. doi: 10.1080/03075070600572090
- Panadero, E., and Broadbent, J. (2018). "Developing evaluative judgment: self-regulated learning perspective," in *Developing Evaluative Judgement in Higher Education. Assessment for Knowing and Producing Quality Work*, eds D. Boud, R. Ajjawi, P. Dawson, and J. Tai (London: Routledge), 81–89.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Faculty Ethics Review Board (FERB) of the Faculty of Social and Behavioural Sciences of Utrecht University. Since this research project is a simple observational study that does not involve any interventions but just comprises interviews, the study has not been subject to review by an ethical committee. This is also in accordance with the recommendations of the FERB. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.



- Panadero, E., and Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assess. Educ.* 21, 133–148. doi: 10.1080/0969594X.2013.877872
- Parr, J. M., and Limbrick, L. (2010). Contextualising practice: hallmarks of effective teachers of writing. *Teach. Teach. Educ.* 26, 583–590. doi: 10.1016/j.tate.2009.09.004
- PASS position paper (2012). *The Case for Programme Focused Assessment*. Available online at: [www.pass.brad.ac.uk](http://www.pass.brad.ac.uk)
- Price, M., Carroll, J., O'Donovan, B., and Rust, C. (2011). If I was going there I wouldn't start from here: a critical commentary on current assessment practice. *Assess. Eval. High. Educ.* 36, 479–492. doi: 10.1080/02602930903512883
- Rust, C., Burkill, S., and Whitfield, R. (2012). *Programme Assessment Strategies. International Perspectives. Examples from outside the UK*. Available online at: [www.pass.brad.ac.uk](http://www.pass.brad.ac.uk)
- Rust, C., Price, M., and O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assess. Eval. High. Educ.* 28, 147–164. doi: 10.1080/02602930301671
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instruct. Sci.* 18, 119–144. doi: 10.1007/BF00117714
- Sadler, R. (2009). "Transforming holistic assessment and grading into a vehicle for complex learning," in *Assessment, Learning and Judgement in Higher Education*, ed G. Joughin (Dordrecht: Springer), 45–63.
- 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.* 14, 281–294. doi: 10.1080/09695940701591867
- van der Vleuten, C. P., Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K. J., et al. (2012). A model for programmatic assessment fit for purpose. *Med. Teach.* 34, 205–214. doi: 10.3109/0142159X.2012.652239
- Weigle, S. C. (2002). *Assessing Writing*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511732997
- Wiliam, D. (2011). What is assessment for learning? *Stud. Educ. Eval.* 37, 3–14. doi: 10.1016/j.stueduc.2011.03.001
- Willis, J. (2011). Affiliation, autonomy and assessment for learning. *Assess. Educ.* 18, 399–415. doi: 10.1080/0969594X.2011.604305
- Yin, R. K. (2014). *Case Study Research Design and Methods*. Los Angeles: Sage.
- Zimmerman, B. J. (2000). "Attainment of self-regulation: a social cognitive perspective," in *Handbook of self-regulation*, eds M. Boekaerts, P. R. Pintrich, and M. Zeidner (San Diego, CA: Academic Press), 13–39. doi: 10.1016/B978-012109890-2/50031-7
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: an overview. *Theory Pract.* 41, 64–70. doi: 10.1207/s15430421tip4102\_2

**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 Baartman and Prins. 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.



# Transparency Isn't Spoon-Feeding: How a Transformative Approach to the Use of Explicit Assessment Criteria Can Support Student Self-Regulation

Kieran Balloo<sup>1\*</sup>, Carol Evans<sup>2</sup>, Annie Hughes<sup>3</sup>, Xiaotong Zhu<sup>2</sup> and Naomi Winstone<sup>1</sup>

<sup>1</sup> Department of Higher Education, University of Surrey, Guildford, United Kingdom, <sup>2</sup> Southampton Education School, University of Southampton, Southampton, United Kingdom, <sup>3</sup> Learning and Teaching Enhancement Centre, Directorate for Student Achievement, Kingston University, Kingston upon Thames, United Kingdom

## OPEN ACCESS

### Edited by:

Anders Jönsson,  
Kristianstad University, Sweden

### Reviewed by:

Joanna Hong-Meng Tai,  
Deakin University, Australia  
Catarina Andersson,  
Umeå University, Sweden

### \*Correspondence:

Kieran Balloo  
k.balloo@surrey.ac.uk

### Specialty section:

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

**Received:** 15 May 2018

**Accepted:** 09 August 2018

**Published:** 03 September 2018

### Citation:

Balloo K, Evans C, Hughes A, Zhu X  
and Winstone N (2018) Transparency  
Isn't Spoon-Feeding: How a  
Transformative Approach to the Use  
of Explicit Assessment Criteria Can  
Support Student Self-Regulation.  
Front. Educ. 3:69.  
doi: 10.3389/feduc.2018.00069

If little care is taken when establishing clear assessment requirements, there is the potential for spoon-feeding. However, in this conceptual article we argue that transparency in assessment is essential to providing equality of opportunity and promoting students' self-regulatory capacity. We begin by showing how a research-informed inclusive pedagogy, the *EAT Framework*, can be used to improve assessment practices to ensure that the purposes, processes, and requirements of assessment are clear and explicit to students. The *EAT Framework* foregrounds how students' and teachers' conceptions of learning (i.e., whether one has a transactional or transformative conception of learning within a specific context) impact assessment practices. In this article, we highlight the importance of being explicit in promoting access to learning, and in referencing the *EAT Framework*, the importance of developing transformative rather than transactional approaches to being explicit. Firstly, we discuss how transparency in the assessment process could lead to "criteria compliance" (Torrance, 2007, p. 282) and learner instrumentalism if a transactional approach to transparency, involving high external regulation, is used. Importantly, we highlight how explicit assessment criteria can hinder learner autonomy if paired with an overreliance on criteria-focused 'coaching' from teachers. We then address how 'being explicit with assessment' does not constitute spoon-feeding when used to promote understanding of assessment practices, and the application of deeper approaches to learning as an integral component of an inclusive learning environment. We then provide evidence on how explicit assessment criteria allow students to self-assess as part of self-regulation, noting that explicit criteria may be more effective when drawing on a transformative approach to transparency, which acknowledges the importance of transparent and mutual student-teacher communications about assessment requirements. We conclude by providing recommendations to teachers

and students about how explicit assessment criteria can be used to improve students' learning. Through an emphasis on transparency of process, clarity of roles, and explication of what constitutes quality within a specific discipline, underpinned by a transformative approach, students and teachers should be better equipped to self-manage their own learning and teaching.

**Keywords:** assessment, feedback, criteria, higher education, inclusive curriculum, self-regulation, spoon-feeding, transparency

## INTRODUCTION

A fundamental goal of higher education has to be to support learners to manage their own learning for themselves both in the present, and in the future as part of sustainable learning practices (Boud, 2000; Boud and Soler, 2016); all aspects of the assessment process should support this (Evans, 2016). In order to increase the effectiveness of assessment in higher education, it has been proposed that assessment should be a learning opportunity that directs students' focus toward what should be learned and engages them in the learning process (Boud and Associates, 2010). Explicit introduction, induction, and appropriate on-going support for the contextual requirements and purposes of learning activities within higher education are therefore important in supporting students' self-regulatory development (Waring and Evans, 2015). However, while students can (arguably) escape from the effects of poor teaching practice, they cannot escape the effects of poorly designed assessment (Boud, 1995a). Assessment practices need to keep pace with twenty-first century learning requirements, and at the same time, be cognizant of the differing contexts, expectations, and needs of our increasingly diverse student body (Balloo, 2017; Balloo et al., 2017).

If little care is taken when establishing clear assessment requirements, there is the potential for "spoon-feeding," yet the move toward transparency in assessment in higher education has largely been positively received (Carless, 2015), since explicit requirements are likely to facilitate fairness in marking practices by enhancing markers' abilities to be consistent in making accurate judgments of student work (Broadbent et al., 2018) and communicating reasons for a particular judgment (Sadler, 2005). Explicit assessment criteria can support students to consider what they are aiming for and how this can be achieved from the perspective of a marker (Nicol and MacFarlane-Dick, 2006), so their learning outcomes move beyond a purely cognitive product, to the development of metacognition (Frederiksen and Collins, 1989; Shephard, 2000; Swaffield, 2011) and assessment literacy (Price et al., 2012). In this article, we present a conceptual analysis of the value of explicit assessment criteria; we highlight the potential risk of spoon-feeding in promoting "criteria compliance" (Torrance, 2007, p. 282), and then we present approaches demonstrating that a careful use of transparency through explicit assessment criteria is crucial to promoting equality of opportunity and students' self-regulation.

## NOTIONS OF "EXPLICIT" WITHIN HIGHER EDUCATION ASSESSMENT PRACTICES

In exploring notions of "explicit" within higher education assessment practices, it is important to consider how students' and teachers' different conceptions of learning (Entwistle and Peterson, 2004) impact on how we enact notions of "explicit" in practice. Notions of being explicit have been covered extensively in the literature, and making assessment processes transparent has a strong history with significant work being undertaken by the Assessment Reform Group (Broadfoot et al., 1999), and notably by Black and Wiliam (1998) in their seminal work on assessment *for* learning. Hattie's "visible learning approach" (Hattie, 2012; Hattie and Yates, 2014) also emphasizes the importance of assessment being explicit. The key issue, however, remains on how being explicit is interpreted and this is where conceptions of learning are central in drawing on our epistemological and ontological assumptions about learning, and our own responsibility in the assessment learning process.

The term "explicit" is a loaded term in relation to how clear information is, and to whom, from an inclusive and critical pedagogy perspective (Waring and Evans, 2015). What is explicit in one context may not be in another; the same student may struggle to grasp meanings from one module to another, timing (in relation to the accumulation of experience and expertise) impacts understandings, and cultural differences implicit in environments and through individual differences impact student and teacher<sup>1</sup> understandings of the learning and teaching context. In addressing student access to learning, a number of key themes emerge from the literature to include the nature and role of scaffolding to support learner understandings, and how this is extended to discussions concerning the accessibility of information, and pedagogical lessons that can be learned from this.

Transparency, clarity, and explicit instruction in assessment have critical roles to play in addressing the long standing and assiduous differentials in student learning outcomes across various student groups. In particular, we argue that assessments that are loosely constructed and lack clarity have the potential to disproportionately disadvantage certain groups of students, and notably, those who have often been referred to as "non-traditional," including those who are first generation

<sup>1</sup>For the sake of brevity and consistency, the term 'teacher' has been used throughout this article to refer to all types of teaching staff in higher education (i.e., educators, academics, lecturers, tutors, professors, etc.).

in higher education, mature learners, students from Black and Minority Ethnic (BME)<sup>2</sup> backgrounds, and those from lower socio-economic personal histories (Newbold et al., 2010). Differential attainment based on socio-economic background and various demographic characteristics is a long standing concern in higher education internationally (HEFCE, 2015; Cahalan et al., 2017; ECU, 2017). Existing research has told us that many students from “non-traditional” backgrounds feel relatively unprepared for the university experience and lack the sense of entitlement held by their white, middle class counterparts (Thomas and Quinn, 2007; Reay et al., 2010). These students are often less conversant with academic language, cultures and traditions, and they lack the confidence to question and challenge normative assessment practices (Southall et al., 2016; Witkowski et al., 2016). With increasing numbers of students engaging in higher education globally, and from increasingly diverse backgrounds, higher education has a significant responsibility to ensure all students have equal access to learning environments. As identified in the “*Feedback Landscape*” (Evans, 2013), a conceptual framework exploring the learning process and individual development from both student and lecturer perspectives through an assessment lens, there are a myriad of individual difference variables impacting a student’s learning. The design of the learning environment may inadvertently advantage some students over others; in understanding the principles of universal design (initiated by Ron Mace at the North Carolina State University College of Design, USA, and initially applied to architecture and then more widely to inclusive pedagogies, Rogers-Shaw et al., 2017), it is important to provide *adaptive* (i.e., all learners can access the learning environment) rather than *adapted* (i.e., learning environments designed to suit a specific type of learner) learning environments (Choi et al., 2009).

*Evans’ Assessment Tool (EAT)* (Evans, 2016) is pertinent to discussions of transparency and inclusivity in assessment practices. Based on a comprehensive synthesis of the assessment feedback literature in higher education (Evans, 2013), the *EAT Framework* was developed to provide research-informed guidance at student, teacher, program, and institution level across three core dimensions of assessment practice: Assessment literacy, assessment feedback, and assessment design. The *EAT Framework* promotes a transformative approach to learning through enactment of its underpinning principles that promote student ownership and autonomy in learning as part of a self-regulated approach to learning. The framework emphasizes the development of student self-regulation, encompassing metacognitive, cognitive, and affective elements (Vermunt and Verloop, 1999). Conceptions of learning impact teaching and learning (Pedrosa-de-Jesus and da Silva Lopes, 2011), and especially the delivery and interpretation of assessment guidance. Being explicit about assessment, and how this is understood, will depend very much on whether one has a transactional or transformative conception of learning within a specific context; the former seeing learning as acquiring, gifting, acting on, and the latter seeing learning as focusing on abstraction of ideas,

ownership, and adaptation of ideas to support understanding and application of learning (Säljö, 1979; Marton et al., 1993). **Table 1**, drawing on the core dimensions underpinning the *EAT Framework*, highlights the importance of student engagement in all decisions around assessment practices as part of developing agency, ownership, and importantly, “*knower-ship*” of the requirements of the discipline (Evans, 2018).

In the *EAT Framework*, the importance of being explicit in relation to higher education academic assessment practices is made within the context of promoting student self-regulation and independence in learning (e.g., consideration of the requirements of tasks and how they are assessed; the roles of all those involved in the assessment process, approaches used and tools available; and in addressing self-management of assessment through open dialogue about what is problematic and uncomfortable in learning). In this context, the moderators (individual and environmental) impacting student access to learning are paramount. What is seen as explicit to some, will not be to others, given learners’ and teachers’ similar and different frames of reference, experiences, and prior knowledge to mention just a few of the variables concerned in this complex equation. **Table 1** illustrates how conceptions of learning impact dimensions of the assessment process using the *EAT Framework*, but also noting the importance of the interaction of context and the demands of the task where in certain circumstances, a transactional approach may be the most appropriate. This suggests the need for a flexible approach which acknowledges the need to “[abandon] the rigid explicit instruction versus minimal guidance dichotomy and [replace] it with a more flexible approach” (Kalyuga and Singh, 2016, p. 833) taking into account the requirements of the task, while being attuned to the needs of students (at the group and individual level). We will now discuss transactional approaches to being transparent with assessment practices, highlighting situations in which these approaches could risk spoon-feeding students.

## AT RISK OF SPOON-FEEDING? A TRANSACTIONAL APPROACH TO TRANSPARENCY IN ASSESSMENT PRACTICES

Transparency in assessment can lead to learner instrumentalism and students having an increased dependence on teachers, since it may be interpreted in a transactional way that sees assessment as something done *to* rather than *with* students (i.e., providing coaching, reading drafts, multiple opportunities for practice, etc.) (Torrance, 2007), with students taking very little ownership of the process. Sadler (2007) notes how criteria have the effect of breaking down assessments into “pea-sized bits to be swallowed one at a time” (p. 390), and that coaching has been utilized as a way to get students to address outcomes rather than actually learn. In this sense, explicit criteria and learning objectives run contrary to the spirit of higher education: “Many people learned many things long before the language of ‘learning goals’ was invented.” (Torrance, 2012, p. 331). From a student perspective, those who favor and/or have been inducted into transactional approaches to learning where external regulation of learning has

<sup>2</sup>Black and Minority Ethnic (BME) is the terminology usually used in the UK to refer to individuals from a non-white background.

**TABLE 1** | Evans (2018) Transformative approaches to assessment practices using the *EAT Framework* compared to transactional approaches.

| Dimension of assessment practice | Transactional approach   | Transformative approach   |
|----------------------------------|--|---|
| Assessment literacy              | <ul style="list-style-type: none"> <li>• Telling—one directional guidance from lecturer to student. Student receives as a gift.</li> <li>• Teacher driven rubrics.</li> <li>• Provision of exemplars.</li> <li>• Provision of assessment criteria.</li> <li>• Provision of glossaries.</li> <li>• Provision of information on assessment regulations.</li> </ul>   | <ul style="list-style-type: none"> <li>• Explaining/discussing requirements with students. Student interacts with information.</li> <li>• Student generated rubrics.</li> <li>• Dialogue around exemplars. Student development of exemplars; students unpacking examples.</li> <li>• Students working with assessment criteria reshaping it in their own language.</li> <li>• Student-/teacher-generated glossaries.</li> <li>• Interpretation of what regulations mean; students contributing to policy.</li> </ul>  |
| Assessment feedback              | <ul style="list-style-type: none"> <li>• Reliance on the teacher for feedback.</li> <li>• Corrective feedback—one directional from teacher to student.</li> <li>• Provision of guidance on how to improve.</li> <li>• Asks students to reflect on their feedback.</li> <li>• Directive. Solutions provided.</li> <li>• Focus on the immediate requirements of the module task.</li> </ul>  | <ul style="list-style-type: none"> <li>• Reliance on range of sources for feedback including self-feedback—emphasis on supporting students to self-assess accurately.</li> <li>• Provision of examples of how to correct with the responsibility on the student to apply the approach throughout their work.</li> <li>• Feedback focused on key areas that need development in relation to student starting points. Student responsibility for developing action plan based on feedback.</li> <li>• Explores with students their interpretation of feedback and self-remediating strategies, and provides models/tools in order to support understandings.</li> <li>• Challenging the student to find solutions.</li> <li>• Problems made explicit. Uncomfortableness of learning and difficulties in managing feedback articulated.</li> <li>• Focus on how all assessment elements fit together within a programme, and application of learning within and beyond the programme.</li> </ul>   |
| Assessment Design                | <ul style="list-style-type: none"> <li>• Assessment tasks designed <i>for</i> students.</li> <li>• Summative assessment by teachers.</li> <li>• Lecturer ownership of assessment tasks.</li> <li>• No choice in assessments.</li> <li>• Tasks designed exclusively to meet specific learning outcomes.</li> <li>• Strongly scaffolded learning tasks—high external regulation.</li> <li>• Resources to support learning provided, but links between them not explicit.</li> <li>• Guidance mainly provided from within by module team.</li> <li>• Resources provided for students.</li> <li>• Limited opportunities for self-assessment.</li> <li>• Limited opportunities to explore assessment holistically and to explore potential issues. Teacher directs solution-finding.</li> </ul> | <ul style="list-style-type: none"> <li>• Assessment tasks designed <i>with</i> students as part of a co-construction model.</li> <li>• Self and teacher summative assessment.</li> <li>• Student/teacher ownership of assessment tasks.</li> <li>• Lecturer mediated choice in assessments.</li> <li>• Tasks designed to meet learning outcomes and to go beyond.</li> <li>• Early scaffolding and subsequent removal of scaffolding to support independence in learning; student awareness of scaffolding needed.</li> <li>• All key resources available from the outset to enable student control of learning but signposted in relation to tasks and key crunch points identified with students. Clear links to resources provided in one place.</li> <li>• Students supported to build networks of support within and beyond module.</li> <li>• Students/teachers responsible for the generation of resources.</li> <li>• Ongoing aligned opportunities for self-assessment.</li> <li>• Key threshold concepts identified from the outset. Students encouraged to provide resources to support understanding in areas seen as difficult and to find own solutions.</li> </ul> |

been high, may want and need explicit guidance, whereas those more used to self-regulating their own learning, may value and also request explicit guidance to a lesser extent (Bell et al., 2013). In such contexts, a vicious circle can be set up where for some groups of students, the provision of more and more guidance may not ever be enough. For example, those with lower levels of self-regulation may be highly dependent on feedback (Çakir et al., 2016), but they may also be less able to use it well.

However, if students are not aware of the standards required of them, misunderstandings about “what constitutes good” can occur (Gibbs and Simpson, 2004). These misunderstandings will then require additional support from teachers to clarify the unclear expectations, which diverts attention away from actual learning and has the potential to disadvantage some

student groups, as identified earlier. For example, where an assessment brief<sup>3</sup> lacks clarity, the normative expectation held by teachers is that their students discuss with each other the requirements of the assessment. There are also situations (mostly those that require creative and original contributions) in which teachers need to draw on their tacit knowledge of what constitutes quality in that domain, so they are not able to explicitly state all of the criteria upfront and thus criteria may remain “fuzzy” to students (Sadler, 1987, 1989). Explicit criteria and learning outcomes alone may therefore

<sup>3</sup>An assessment brief is a document that states the purpose of an assessment and provides a clear explanation of what is expected of students (Gilbert and Maguire, 2014).



not be able to convey teachers' tacit knowledge (O'Donovan et al., 2004), so students may be dependent on the teacher until they have enough understanding of how to interpret and access this knowledge (Sadler, 1989). There is clearly an argument for allowing teachers the flexibility to make qualitative human judgments, yet we argue that assessments requiring additional engagement between students and their teachers, or indeed between students themselves, to clarify expectations and requirements, are not innately inclusive. This culture is likely to further create exclusivity in students' opportunities to access assessment guidance; not all students will understand the manner in which this can adequately be obtained, particularly students from non-traditional backgrounds who are less willing to seek advice and guidance (Francis, 2008).

For many commentators, the riposte to the issue of differential attainment centers around the concept of the inclusive curriculum (Berry and Loke, 2011; Singh, 2011; Stevenson, 2012). An inclusive curriculum in higher education is one designed and delivered to engage students in learning that is accessible, relevant and meaningful to students from a wide range of backgrounds (Hockings, 2010). Following this logic, inclusive assessment should be accessible, applicable, expressive and clearly communicated. The principles of inclusive assessment have been expressed through the need to offer a varied diet of assessment types; the argument here being that a diverse "mix" of assessment methods will ensure that students with certain skills are not disadvantaged by specific forms of assessment. However, while giving choice in assessment method can have a positive effect on students who have clear understandings of their strengths and weaknesses by empowering them and allowing them to take responsibility for their learning, choice can also act to disempower and overwhelm students. Furthermore, an overemphasis on choice for choice's sake takes away from careful consideration of what the most appropriate assessment tasks are to enable a student to best meet required learning outcomes. We argue that the most inclusive assessment practices are ones that support and consciously scaffold students' learning through the underpinning of good assessment design.

Early assessment tasks may need to be more strongly scaffolded than later tasks, for example, in the provision of detailed explicit assessment criteria, to ensure equality of opportunity. However, if there is no gradual removal of this scaffolding as students gain more experience, this runs the risk of spoon-feeding, and increasing student dependence rather than independence in learning. Spoon-feeding in education can be defined as the process of teachers directly telling students everything they need to know about the requirements of a specific task, thus requiring little independent thought on their part (Smith, 2008). Epistemologically, spoon-feeding could be viewed as stemming from a representational model in which teachers merely transmit knowledge to passive students (Raelin, 2009) who have been socialized into a culture of dependence on them (Dehler and Welsh, 2014). "A complaint I hear from [university teachers] is that, undergraduate students require 'spoon-feeding'.... They say that their students demand it, and feel that they must unwillingly oblige. The metaphor of spoon-feeding doesn't match their idea of what a teacher should do, or how students should

be going about their learning. They complain that students just want to be told exactly what to do, the facts, the right answers, instead of thinking things through for themselves." (Smith, 2008, p. 715).

In the context of assessment, spoon-feeding may involve explicitly telling students *what* they need to do for an assignment, and *how* to meet the assessment criteria, without leaving it up to them to ascertain this for themselves. Addressing task criteria in the absence of understanding the domain being assessed has been termed by Torrance (2007) as "criteria compliance" (p. 282). Some students may use explicit criteria to focus on exactly what needs to be done to reach a desired level of achievement, rather than actually learning material fully (Panadero and Jonsson, 2013). Students' and teachers' conceptions of learning play a role in this; if teachers simply supply assessment requirements to students in a transactional manner, so they can passively "check boxes," it is unlikely that students will engage with the criteria in a way that will develop their learning and self-regulation.

Nonetheless, explicit assessment criteria can directly pave the way for self-regulation to occur. Evidence has shown that explicit criteria have a positive effect on all phases of the self-regulation process<sup>4</sup> (Panadero and Romero, 2014). For example, criterion-referencing<sup>5</sup> is a common characteristic of self-assessment<sup>6</sup> (Andrade and Du, 2007), which can be seen as a form of self-feedback (Andrade and Du, 2007; Winstone et al., 2017) encompassing the self-regulatory skills of self-monitoring and self-evaluation (Panadero et al., 2017). Self-assessment can foster self-efficacy, motivation to learn, and in turn, superior performance (Schunk, 2003; Andrade and Valtcheva, 2009). One way to facilitate self-assessment is through the use of rubrics<sup>7</sup> (Panadero and Romero, 2014). A rubric, by definition, needs to make use of explicit assessment criteria (Jones et al., 2017) and Popham (1997) notes how a set of evaluative criteria is the most important aspect of a rubric, because mastery over these criteria will eventually result in skill mastery. The transparency provided by rubrics lays the groundwork for feedback to be interpreted; students' expectations are clarified, their attention is more closely focused on what their assessments require of them, they gain greater perceived control and confidence about their assessments, and their anxiety about completing the assessment is reduced (Andrade and Du, 2005; Andrade and Valtcheva, 2009; Panadero and Jonsson, 2013; Jonsson, 2014). Self-assessment affords students the opportunity to receive feedback that they are likely to perceive as low- or no-stakes when compared to teacher feedback (Chen et al., 2017). Furthermore, Panadero

<sup>4</sup>Zimmerman's 2002 model of self-regulated learning proposes that there are three phases encompassing the following: a forethought phase, which occurs before learning and involves planning and goal setting; a performance phase, which takes place during learning and involves self-monitoring, through which students track their progress toward goals; and a self-reflection phase, which happens after learning and involves self-evaluation in which students judge their performance against a set of standards.

<sup>5</sup>Criterion-referencing involves the use of clear assessment criteria to determine whether specific learning outcomes have been met (Torrance, 2007).

<sup>6</sup>Self-assessment is the act of posing questions to oneself in order to make judgments about whether certain criteria and standards are being met (Boud, 1995b).

<sup>7</sup>Rubrics are written documents that communicate the criteria of an assessment and the levels of quality expected (Andrade, 2000).

et al. (2013) found that rubrics reduced students' use of negative self-regulatory actions (i.e., self-regulatory approaches that are motivated by a desire to endorse performance avoidance goals, such as trying to avoid failing). Thus, a clear understanding of explicit standards and criteria serves as a crucial prerequisite to engaging in activities that enhance self-regulation (Andrade and Valtcheva, 2009).

However, the presence of explicit criteria alone does not mean they will automatically be used by students to self-regulate. Through focus group discussions with students, Andrade and Du (2007) reported evidence of tensions between what students thought was required of them in their self-assessment, and teachers' actual expectations of their work. For example, one student espoused that self-assessment was really just assessing what the teacher wanted, since they were the ones who set the assessment criteria. If students only use self-assessment to determine how to please the teacher and not internalize standards, it is hard to see how self-assessment might foster autonomy in students' future approaches to assessment. Similarly, Handley and Williams (2011) found that students did not understand how exemplar work related to explicit assessment criteria unless teachers directly showed them how this work mapped onto the criteria. Since rubrics make clear to students how their work will be evaluated and graded, there is a perceived fairness to using them (Reddy and Andrade, 2010), which is likely why students find them to be desirable, even in the absence of understanding the meaning of the criteria or how to apply them (Jonsson, 2014). A corollary of this is that students need to have explicit criteria in order to self-assess, but the manner in which these criteria are established can be done in a transactional or transformative way. For example, Fraile et al. (2017) claimed that involving students in the co-creation of criteria can counter any notion that self-assessment using rubrics could hinder their autonomy. Therefore, in order to avoid spoon-feeding, teachers should consider moving beyond the transactional approach of simply providing students with explicit criteria. A transformative approach acknowledges the roles of both teachers and students in assessment, which maximizes opportunities for enhancing students' self-regulatory capacities. Therefore, we need to carefully consider how we use assessment tools to support student independence rather than dependence in learning so that students take charge of these tools, make them their own, and use them appropriately. In doing so, students are able to demonstrate understanding of the requirements of learning.

## BEYOND SPOON-FEEDING: A TRANSFORMATIVE APPROACH TO TRANSPARENCY IN ASSESSMENT PRACTICES

Developing students' assessment literacies is one of the most effective ways to address differential attainment and improve students' learning (Price et al., 2012). Providing explicit

assessment criteria and rubrics is important, but there is also an unequivocal requirement to explore the assessment *with* students in timetabled classroom sessions to ensure that students get the opportunity to speak to their teacher(s) and their peers to clarify misconceptions and solidify expectations (Bloxham and West, 2004). Classroom interventions where students can unpick assessment criteria and rubrics, then rewrite them in their own words, are particularly successful for students who are less confident to approach teachers in their office hours or after class. These interventions are also important for the ever-increasing numbers of students who commute to university (Thomas and Jones, 2017), and for those students whose face-to-face engagement with their peers and teachers may be more limited. Additional strategies include involving students in assessing previous work and articulating why they received the mark that they did, and peer marking formative work. Taras (2001), drawing on Sadler's work, carried out an innovative intervention in which students had the opportunity to propose their own criteria before comparing these to the actual criteria that had been set. Students' work was then returned without a grade and they needed to self-assess their work against agreed criteria based on the feedback they had received. Teachers believed these approaches allowed students who failed to understand why, and students felt that it made them more aware of what their assessment requirements actually were. Similarly, Evans and Waring (2011) suggested that students' engagement with assessment tasks could be deeper and more independent where clear assessment criteria had been determined through dialogue between students and teachers. They found that, during initial teacher education, student teachers valued clarity in assessment requirements, because it allowed them to plan their work effectively and complete it under tight time constraints. The important aspect here is shared understandings between teachers and students of what criteria mean within the specific context of a task. As **Table 1** shows, a transformative approach emphasizes elements of practice that give students ownership over the assessment process; assessment is something done *with* rather than *to* students.

The provision of explicit guidance is aligned to Vygotskian notions of the *zone of proximal development* (ZPD) (Vygotsky, 1978), involving learning support from a knowledgeable "other" in order to make progress; this "other" could include peers, friends, networks, media, internet, journals, etc. The main issue with the ZPD is how to take learning to another level; as what and who can support the achievement of this are critical. Using the *EAT Framework*, instruction is centered on supporting students to be more proactive in attaining this support for themselves, but this does not preclude the role of the teacher in this endeavor (see also Nash and Winstone, 2017). A principal aim of "being explicit" is to enable students to focus on the elements of learning that are most important, rather than becoming embedded in the minutiae. The level of scaffolding is very much dependent on students' individual differences, timing (i.e., the stage of development of the learner within the learning process, such as novice or expert, Neubrand et al., 2016), the nature of the task (Arnold et al., 2014), and

alignment of the level of complexity of information with the expertise of the learner (Vogel-Walcutt et al., 2011). From a pedagogical perspective, getting the level of scaffolding correct is crucial; this also includes the removal/fading of such support in order to support learners to be better able to transfer abilities from one context to another (Fang et al., 2016; Yuriev et al., 2017). We know that too much scaffolding can lead to student dependence rather than independence in learning (Koopman et al., 2011). Within learning contexts, there needs to be sufficient challenge (constructive friction) to support learning as opposed to destructive friction (where the learning context is too overwhelming) (Silén and Uhlin, 2008). Furthermore, as noted by Blasco (2015), some disruption or level of discomfort is often needed in order for students to be able to use explicit guidance, experiment, and then be able to integrate concepts and ideas into their own knowledge structures in order to be able to use within their own contexts.

A key part of scaffolding is supporting student access to information, networks and resources; fundamental to this is an understanding that individuals process information in different ways (Kozhevnikov et al., 2014). How information is presented impacts an individual's cognitive and emotional regulation in terms of the impact it has on their cognitive load (van Merriënboer and Sweller, 2005), which is also affected by the emotional meanings attached to specific verbal and visual representations. Cognitive load theory (CLT) is based on the assumption that our working memory capacity (i.e., our ability to temporally hold, while concurrently processing, information) is limited (Howard-Jones, 2010), and we therefore need to consider how we can either increase our working memory capacity, and/or reduce load to facilitate student access to information and subsequent recall. Key pedagogical lessons are the need to ensure information is presented in the most appropriate way for the requirements of the task and that any potentially distracting information is removed. Presenting information in visual and verbal forms aligned to visual and verbal processing systems drawing on dual coding theory has also been found to be successful in enhancing memory (Paivio, 2006). CLT especially reminds us of the importance of not overloading students with information at a time when they may not be in a position to process it (e.g., overloading with assessment information at the start of a module when this will not be a priority in relation to managing more immediate tasks).

The need for being explicit in clarifying students' "*understanding of good*" is a necessity, in that if we do not have a clear idea of what we are aiming for, it is hard to get there, albeit not impossible (Ramaprasad, 1983; Sadler, 2010). Evans (2016), in her pragmatic articulation of the assessment literature and extensive work in higher education practice using the *EAT Framework* as part of "*feedback exchange*" (Evans, 2013), argues for the importance of shared understandings between lecturers and students about not only "what constitutes good," but about conceptions of learning in the first place, as this underpins how the notion of being explicit is enacted. The notion of feedback exchange is critical as part of this equation, as it extends the dialogue beyond the immediate teacher-student relationship to

consider all the available information that students can access from a range of sources. The critical issue is with assessment design and training; providing affordances and supporting students in being able to maximize support from a range of sources within and beyond the immediate learning environment is imperative (Evans, 2013). The formative assessment literature is also germane to discussions about the importance of transparent and mutual student-teacher communications (Scott et al., 2014) in terms of ensuring that the two parties share common understandings about the nature of assessment tasks, what the learning process entails, and how evidence of learning is being assessed. Black and Wiliam (1998) initially emphasized that one of the key facets of formative assessment involved teachers sharing success criteria with their students. However, subsequent elucidation of the strategies involved in formative assessment now highlights the shared student and teacher roles in this process; alongside teachers having the responsibility to clarify criteria, students also have the responsibility to understand these criteria (Black and Wiliam, 2009). The main aim of formative assessment is now seen to be the development of self-regulation (Panadero et al., 2018). Thus, changing the goals of formative assessment practices can facilitate a move from transactional to transformative approaches.

Whether explicit guidance on assessment criteria and the assessment process gives students entry to the nature of knowledge within a discipline and its requirements is debatable. In trying to be explicit with a transformative approach, we are aiming to improve students' "*knower-ship*" of a subject or context(s), but this takes interaction between disciplinary insiders and students to come to shared understandings of what disciplines want their students to become, to know, and how they want students to construct knowledge (van Heerden et al., 2017). Richards and Pilcher (2014) highlight the importance of shared negotiation of meanings as part of teacher-student dialogue in their promotion of an "anti-glossary" approach. They argue that all terms are loaded (disciplinary, cultural, temporal and spatial inferences) and have different meanings for different actors, and for the same actors in different contexts, and over time; therefore, in order for a glossary of key terms to be usable, it does need to be deconstructed and reformulated through the medium of dialogue. Where teachers hold their own tacit knowledge of a domain that cannot easily be articulated as explicit assessment criteria, discussions between teachers and students about explicit criteria and standards can lead to the student forming their own tacit understanding of quality judgments (Yucel et al., 2014). Attention needs to be focused on making the implicit explicit (transparency in all higher education processes and disciplinary norms), and in attending to student dispositions (McCune and Entwistle, 2011), so that they are in a position to make the most of affordances within the learning environment. Going beyond the written to ensuring dialogic approaches to support shared understandings of what is required, is also highlighted by Papadopoulos et al. (2013) in their work with students on techniques in assessment to support students by being explicit, and also in Carless and Chan's (2017) work on the dialogic use of exemplars by teachers with students.

## CONCLUSIONS

A key aim of higher education has to be to support learners to become more independent in their learning; ensuring them access to learning through being explicit is essential to this endeavor. The *EAT Framework* provides one example where, through a holistic approach, all aspects of assessment practice (promotion of assessment literacy, assessment feedback, and assessment design) are underpinned by the need to support students in managing learning for themselves. In this article, we have discussed the different ways in which explicit standards, criteria, tools, and processes can lead to differential impacts on students, with much depending on how “explicit” is enacted and received. However, we need to be mindful of individual differences in learners’ contexts; we may endeavor to have all students and teachers working at a deeper and more transformative level, but this may not always be appropriate, so a flexible approach should be taken. We argue that a transactional approach to the use of explicit assessment criteria may run the risk of spoon-feeding students, so the ultimate goal of assessment in higher education should be to move to a transformative approach, striving for shared understandings between teachers and students of assessment requirements. Thus, the implications and recommendations of this article are also shared between teachers and students. If teachers strongly scaffold early tasks in an effort to be more inclusive, they need to have clear plans for how and when to fade this scaffolding once there is an expectation that students should make more of their own judgments; clarifying role expectations from the outset is a key part of this.

Interventions that give students opportunities to discuss and work with criteria are more likely to be effective in developing students’ self-regulation than the mere provision of criteria alone (e.g., use of rubrics, deconstruction of assessment criteria, etc.). Engaging students in all decisions around assessment practices allows them to develop an understanding of what constitutes

quality within their discipline, so students can become better equipped to self-manage their own learning; this emphasizes the importance of training in assessment for staff and students.

Implementing “explicit” in a robust way, therefore, requires learners and teachers to develop shared conceptions of learning that bring to attention what it is to learn in a meaningful way within a given context as part of a joint endeavor. Teachers’ timetabled classroom sessions should give all students equal opportunities to develop their own understanding of explicit assessment criteria. As part of this, students need to be carefully inducted into their responsibility within assessment if they are to become more self-regulatory in their approach to assessments within and beyond a specific context (Evans, 2013; Nash and Winstone, 2017). The provision of explicit assessment criteria should be seen as the starting point to developing their own understanding of how to address these criteria. The fundamental point here, as advocated in the *EAT Framework*, is that if we see students as co-constructors of the curriculum, there is no reason why, if they are given appropriate training in assessment design, they cannot develop and design criteria for themselves. If this is achieved, a transformative approach to transparency in assessment is far from spoon-feeding.

## AUTHOR CONTRIBUTIONS

All authors made substantial contributions to the preparation, writing and revisions for this article.

## FUNDING

This work was supported by funding from the Office for Students (OfS), England, and University of Southampton, University of Surrey, and Kingston University, through the Maximizing Student Success through the Development of Self-Regulation project award led by CE (grant number L16).

## REFERENCES

- Andrade, H. (2000). Using rubrics to promote thinking and learning. *Educ. Leadersh.* 57, 13–18.
- Andrade, H., and Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Pract. Assess. Res. Eval.* 10, 1–11.
- Andrade, H., and Du, Y. (2007). Student responses to criteria referenced self-assessment. *Assess. Eval. High. Educ.* 32, 159–181. doi: 10.1080/02602930600801928
- Andrade, H., and Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Pract.* 48, 12–19. doi: 10.1080/00405840802577544
- Arnold, J. C., Kremer, K., and Mayer, J. (2014). Understanding students’ experiments—What kind of support do they need in inquiry tasks? *Int. J. Sci. Educ.* 36, 2719–2749. doi: 10.1080/09500693.2014.930209
- Balloo, K. (2017). In-depth profiles of the expectations of undergraduate students commencing university: A Q methodological analysis. *Stud. High. Educ.* 1–12. doi: 10.1080/03075079.2017.1320373
- Balloo, K., Pauli, R., and Worrell, M. (2017). Undergraduates’ personal circumstances, expectations and reasons for attending university. *Stud. High. Educ.* 42, 1373–1384. doi: 10.1080/03075079.2015.1099623
- Bell, A., Mladenovic, R., and Price, M. (2013). Students’ perceptions of the usefulness of marking guides, grade descriptors and annotated exemplars. *Assess. Eval. High. Educ.* 38, 769–788. doi: 10.1080/02602938.2012.714738
- Berry, J., and Loke, G. (2011). *Improving the Degree Attainment of Black and Minority Ethnic Students*. York: Equality Challenge Unit (ECU)/Higher Education Academy (HEA). Available online at: <http://www.ecu.ac.uk/wp-content/uploads/external/improving-degree-attainment-bme.pdf>
- Black, P., and Wiliam, D. (1998). Assessment and classroom learning. *Assess. Educ. Princ. Policy Pract.* 5, 7–74. doi: 10.1080/0969595980050102
- Black, P., and Wiliam, D. (2009). Developing the theory of formative assessment. *Educ. Assess. Eval. Account.* 21, 5–31. doi: 10.1007/s11092-008-9068-5
- Blasco, M. (2015). Making the tacit explicit: rethinking culturally inclusive pedagogy in international student academic adaptation. *Pedagog. Cult. Soc.* 23, 85–106. doi: 10.1080/14681366.2014.922120
- Bloxham, S., and West, A. (2004). Understanding the rules of the game: Marking peer assessment as a medium for developing students’ conceptions of assessment. *Assess. Eval. High. Educ.* 29, 721–733. doi: 10.1080/0260293042000227254
- Boud, D. (1995a). “Assessment and learning: contradictory or complementary?” in *Assessment for Learning in Higher Education*, ed. P. Knight (London: Kogan Page), 35–48.



- Boud, D. (1995b). *Enhancing Learning Through Self-Assessment*. Oxon: Routledge.
- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Stud. Contin. Educ.* 22, 151–167. doi: 10.1080/713695728
- Boud, D., and Associates (2010). *Assessment 2020: Seven Propositions for Assessment Reform in Higher Education*. Sydney, NSW: Australian Learning and Teaching Council.
- Boud, D., and Soler, R. (2016). Sustainable assessment revisited. *Assess. Eval. High. Educ.* 41, 400–413. doi: 10.1080/02602938.2015.1018133
- Broadbent, J., Panadero, E., and Boud, D. (2018). Implementing summative assessment with a formative flavour: a case study in a large class. *Assess. Eval. High. Educ.* 43, 307–322. doi: 10.1080/02602938.2017.1343455
- Broadfoot, P., Daugherty, R., Gardner, J., Gipps, C., Harlen, W., James, M., et al. (1999). *Assessment for Learning: Beyond the Black Box*. Cambridge: Nuffield Foundation and University of Cambridge. Available online at: [http://www.nuffieldfoundation.org/sites/default/files/files/beyond\\_blackbox.pdf](http://www.nuffieldfoundation.org/sites/default/files/files/beyond_blackbox.pdf)
- Cahalan, M., Perna, L. W., Yamashita, M., Ruiz, R., and Franklin, K. (2017). *Indicators of Higher Education Equity in the United States: 2017 Trend Report*. Washington, DC: Pell Institute for the Study of Higher Education, Council for Education Opportunity (COE) and Alliance for Higher Education and Democracy (AHEAD) of the University of Pennsylvania. Available online at: [http://pellinstitute.org/downloads/publications-Indicators\\_of\\_Higher\\_Education\\_Equity\\_in\\_the\\_US\\_2017\\_Historical\\_Trend\\_Report.pdf](http://pellinstitute.org/downloads/publications-Indicators_of_Higher_Education_Equity_in_the_US_2017_Historical_Trend_Report.pdf)
- Çakir, R., Korkmaz, Ö., Bacanak, A., and Arslan, Ö. (2016). An exploration of the relationship between students' preferences for formative feedback and self-regulated learning skills. *Malaysian Online J. Educ. Sci.* 4, 14–30.
- Carless, D. (2015). Exploring learning-oriented assessment processes. *High. Educ.* 69, 963–976. doi: 10.1007/s10734-014-9816-z
- Carless, D., and Chan, K. K. H. (2017). Managing dialogic use of exemplars. *Assess. Eval. High. Educ.* 42, 930–941. doi: 10.1080/02602938.2016.1211246
- Chen, F., Lui, A. M., Andrade, H., Valle, C., and Mir, H. (2017). Criteria-referenced formative assessment in the arts. *Educ. Assess. Eval. Account.* 29, 297–314. doi: 10.1007/s11092-017-9259-z
- Choi, I., Lee, S. J., and Kang, J. (2009). Implementing a case-based e-learning environment in a lecture-oriented anaesthesiology class: do learning styles matter in complex problem solving over time? *Br. J. Educ. Technol.* 40, 933–947. doi: 10.1111/j.1467-8535.2008.00884.x
- Dehler, G. E., and Welsh, M. A. (2014). Against spoon-feeding. For learning. Reflections on students' claims to knowledge. *J. Manag. Educ.* 38, 875–893. doi: 10.1177/1052562913511436
- ECU (2017). *Equality in Higher Education: Students Statistical Report*. London: Equality Challenge Unit (ECU). Available online at: [http://hub.ecu.ac.uk/MembersArea/download-publication.aspx?fileName=Students\\_report\\_2017\\_November2017.pdf&login=Y](http://hub.ecu.ac.uk/MembersArea/download-publication.aspx?fileName=Students_report_2017_November2017.pdf&login=Y)
- Entwistle, N. J., and Peterson, E. R. (2004). Conceptions of learning and knowledge in higher education: relationships with study behaviour and influences of learning environments. *Int. J. Educ. Res.* 41, 407–428. doi: 10.1016/j.ijer.2005.08.009
- Evans, C. (2013). Making sense of assessment feedback in higher education. *Rev. Educ. Res.* 83, 70–120. doi: 10.3102/0034654312474350
- Evans, C. (2016). *Enhancing Assessment Feedback Practice in Higher Education: The EAT Framework*. Southampton: University of Southampton. Available online at: <https://eatframework.org.uk/>
- Evans, C. (2018). *A Transformative Approach to Assessment Practices Using the EAT Framework Presentation. Small Scale Projects in Experimental Innovation*. Catalyst A Funding Higher Education Funding Council for England (Southampton: HEFCE).
- Evans, C., and Waring, M. (2011). Student teacher assessment feedback preferences: the influence of cognitive styles and gender. *Learn. Individ. Differ.* 21, 271–280. doi: 10.1016/j.lindif.2010.11.011
- Fang, S. C., Hsu, Y. S., and Hsu, W. H. (2016). Effects of explicit and implicit prompts on students' inquiry practices in computer-supported learning environments in high school earth science. *Int. J. Sci. Educ.* 38, 1699–1726. doi: 10.1080/09500693.2016.1213458
- Frail, J., Panadero, E., and Pardo, R. (2017). Co-creating rubrics: the effects on self-regulated learning, self-efficacy and performance of establishing assessment criteria with students. *Stud. Educ. Eval.* 53, 69–76. doi: 10.1016/j.stueduc.2017.03.003
- Francis, R. A. (2008). An investigation into the receptivity of undergraduate students to assessment empowerment. *Assess. Eval. High. Educ.* 33, 547–557. doi: 10.1080/02602930701698991
- Frederiksen, J. R., and Collins, A. (1989). A systems approach to educational testing. *Educ. Res.* 18, 27–32. doi: 10.3102/0013189X018009027
- Gibbs, G., and Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learn. Teach. High. Educ.* 1, 3–31. doi: 10.1080/07294360.2010.512631
- Gilbert, F., and Maguire, G. (2014). *Developing Academic Communication in Assignment Briefs to Enhance the Student Experience in Assessment*. Available online at: <https://assignmentbriefdesign.weebly.com/>
- Handley, K., and Williams, L. (2011). From copying to learning: using exemplars to engage students with assessment criteria and feedback. *Assess. Eval. High. Educ.* 36, 95–108. doi: 10.1080/02602930903201669
- Hattie, J. (2012). *Visible Learning for Teachers: Maximizing Impact on Learning*. London; New York, NY: Routledge.
- Hattie, J., and Yates, G. (2014). *Visible Learning and the Science of How We Learn*. London: Routledge.
- HEFCE (2015). *Differences in Degree Outcomes: The Effect of Subject and Student Characteristics*. Bristol: Higher Education Funding Council for England (HEFCE). Available online at: [http://www.hefce.ac.uk/media/HEFCE,2014/Content/Pubs/2015/201521/HEFCE2015\\_21.pdf](http://www.hefce.ac.uk/media/HEFCE,2014/Content/Pubs/2015/201521/HEFCE2015_21.pdf)
- Hockings, C. (2010). *Inclusive Learning and Teaching in Higher Education: a Synthesis of Research*. York: Higher Education Academy (HEA) Available online at: [https://www.heacademy.ac.uk/system/files/inclusive\\_teaching\\_and\\_learning\\_in\\_he\\_synthesis\\_200410\\_0.pdf](https://www.heacademy.ac.uk/system/files/inclusive_teaching_and_learning_in_he_synthesis_200410_0.pdf)
- Howard-Jones, P. (2010). *Introducing Neuroeducational Research*. London; New York, NY: Routledge.
- Jones, L., Allen, B., Dunn, P., and Brooker, L. (2017). Demystifying the rubric: a five-step pedagogy to improve student understanding and utilisation of marking criteria. *High. Educ. Res. Dev.* 36, 129–142. doi: 10.1080/07294360.2016.1177000
- Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assess. Eval. High. Educ.* 39, 840–852. doi: 10.1080/02602938.2013.875117
- Kalyuga, S., and Singh, A.-M. (2016). Rethinking the boundaries of cognitive load theory in complex learning. *Educ. Psychol. Rev.* 28, 831–852. doi: 10.1007/s10648-015-9352-0
- Koopman, M., Den Brok, P., Beijard, D., and Teune, P. (2011). Learning processes of students in pre-vocational secondary education: relations between goal orientations, information processing strategies and development of conceptual knowledge. *Learn. Individ. Differ.* 21, 426–431. doi: 10.1016/j.lindif.2011.01.004
- Kozhevnikov, M., Evans, C., and Kosslyn, S. M. (2014). Cognitive style as environmentally sensitive individual differences in cognition. *Psychol. Sci. Public Interes.* 15, 3–33. doi: 10.1177/1529100614525555
- Marton, F., Dall'Alba, G., and Beaty, E. (1993). Conceptions of learning. *Int. J. Educ. Res.* 19, 277–300.
- McCune, V., and Entwistle, N. (2011). Cultivating the disposition to understand in 21st century university education. *Learn. Individ. Differ.* 21, 303–310. doi: 10.1016/j.lindif.2010.11.017
- Nash, R. A., and Winstone, N. E. (2017). Responsibility-sharing in the giving and receiving of assessment feedback. *Front. Psychol.* 8:1519. doi: 10.3389/fpsyg.2017.01519
- Neubrand, C., Borzikowsky, C., and Harms, U. (2016). Adaptive prompts for learning evolution with worked examples-Highlighting the students between the “novices” and the “experts” in a classroom. *Int. J. Environ. Sci. Educ.* 11, 6774–6795.
- Newbold, J. J., Mehta, S. S., and Forbus, P. (2010). A comparative study between non-traditional and traditional students in terms of their demographics, attitudes, behavior and educational performance. *Int. J. Educ. Res.* 5, 1–24.
- Nicol, D., and MacFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Stud. High. Educ.* 31, 199–218. doi: 10.1080/03075070600572090
- O'Donovan, B., Price, M., and Rust, C. (2004). Know what I mean? Enhancing student understanding of assessment standards and criteria. *Teach. High. Educ.* 9, 325–335. doi: 10.1080/1356251042000216642
- Paivio, A. (2006). *Mind and its Evolution: A Dual Coding Theoretical Approach*. London: Routledge.



- 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., Andrade, H., and Brookhart, S. (2018). Fusing self-regulated learning and formative assessment: A roadmap of where we are, how we got here, and where we are going. *Aust. Educ. Res.* 45, 13–31. doi: 10.1007/s13384-018-0258-y
- Panadero, E., and Jonsson, 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., Jonsson, 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., and Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assess. Educ. Princ. Policy Pract.* 21, 133–148. doi: 10.1080/0969594X.2013.877872
- Papadopoulos, P. M., Demetriadis, S. N., and Weinberger, A. (2013). “Make it explicit!”: Improving collaboration through increase of script coercion. *J. Comput. Assist. Learn.* 29, 383–398. doi: 10.1111/jcal.12014
- Pedrosa-de-Jesus, M. H., and da Silva Lopes, B. (2011). The relationship between teaching and learning conceptions, preferred teaching approaches and questioning practices. *Res. Pap. Educ.* 26, 223–243. doi: 10.1080/02671522.2011.561980
- Popham, W. J. (1997). What's wrong—and what's right—with rubrics. *Educ. Leadersh.* 55, 72–75.
- Price, M., Rust, C., O'Donovan, B., Handley, K., and Bryant, R. (2012). *Assessment Literacy: The Foundation for Improving Student Learning*. Oxford: Oxford Centre for Staff and Learning Development.
- Raelin, J. A. (2009). The practice turn-away: forty years of spoon-feeding in management education. *Manag. Learn.* 40, 401–410. doi: 10.1177/1350507609335850
- Ramaprasad, A. (1983). On the definition of feedback. *Behav. Sci.* 28, 4–13. doi: 10.1002/bs.3830280103
- Reay, D., Crozier, G., and Clayton, J. (2010). “Fitting in” or “standing out”: working-class students in UK higher education. *Br. Educ. Res. J.* 36, 107–124. doi: 10.1080/01411920902878925
- Reddy, Y. M., and Andrade, H. (2010). A review of rubric use in higher education. *Assess. Eval. High. Educ.* 35, 435–448. doi: 10.1080/02602930902862859
- Richards, K., and Pilcher, N. (2014). Contextualising higher education assessment task words with an “anti -glossary” approach. *Int. J. Qual. Stud. Educ.* 27, 604–625. doi: 10.1080/09518398.2013.805443
- Rogers-Shaw, C., Carr-Chellman, D. J., and Choi, J. (2017). Universal design for learning: guidelines for accessible online instruction. *Adult Learn.* 29, 104515951773553. doi: 10.1177/1045159517735530
- Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford Rev. Educ.* 13, 191–209.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instr. Sci.* 18, 119–144. doi: 10.1007/BF00117714
- Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assess. Eval. High. Educ.* 30, 175–194. doi: 10.1080/0260293042000264262
- Sadler, D. R. (2007). Perils in the meticulous specification of goals and assessment criteria. *Assess. Educ. Princ. Policy Pract.* 14, 387–392. doi: 10.1080/09695940701592097
- Sadler, D. R. (2010). Beyond feedback: developing student capability in complex appraisal. *Assess. Eval. High. Educ.* 35, 535–550. doi: 10.1080/02602930903541015
- Säljö, R. (1979). Learning about learning. *High. Educ.* 8, 443–451. doi: 10.1007/BF01680533
- Schunk, D. H. (2003). Self-efficacy for reading and writing: influence of modeling, goal setting, and self-evaluation. *Read. Writ. Q.* 19, 159–172. doi: 10.1080/10573560308219
- Scott, D., Hughes, G., Evans, C., Burke, P. J., Walter, C., and Watson, D. (2014). *Learning Transitions in Higher Education*. London: Palgrave Macmillan. doi: 10.1057/9781137322128
- Shephard, L. A. (2000). The role of assessment in a learning culture. *Educ. Res.* 29, 4–14. doi: 10.3102/0013189X029007004
- Silén, C., and Uhlén, L. (2008). Self-directed learning—a learning issue for students and faculty! *Teach. High. Educ.* 13, 461–475. doi: 10.1080/13562510802169756
- Singh, G. (2011). *Black and Minority Ethnic (BME) Students Participation in Higher Education: Improving Retention and Success—A Synthesis of Research Evidence*. York: Higher Education Academy (HEA). Available online at: [https://www.heacademy.ac.uk/system/files/bme\\_synthesis\\_final.pdf](https://www.heacademy.ac.uk/system/files/bme_synthesis_final.pdf)
- Smith, H. (2008). Spoon-feeding: or how I learned to stop worrying and love the mess. *Teach. High. Educ.* 13, 715–718. doi: 10.1080/13562510802452616
- Southall, J., Wason, H., and Avery, B. (2016). Non-traditional, commuter students and their transition to higher education—a synthesis of recent literature to enhance understanding of their needs. *Student Engagem. Exp. J.* 5, 1–15. doi: 10.7190/seej.v4i1.128
- Stevenson, J. (2012). *Black and Minority Ethnic Student Degree Retention and Attainment Introduction: The BME Degree Attainment Learning and Teaching Summit*. York: Higher Education Academy (HEA). Available online at: [https://www.heacademy.ac.uk/system/files/bme\\_summit\\_final\\_report.pdf](https://www.heacademy.ac.uk/system/files/bme_summit_final_report.pdf)
- Swaffield, S. (2011). Getting to the heart of authentic assessment for learning. *Assess. Educ. Princ. Policy Pract.* 18, 433–449. doi: 10.1080/0969594X.2011.582838
- Taras, M. (2001). The use of tutor feedback and student self-assessment in summative assessment tasks: towards transparency for students and for tutors. *Assess. Eval. High. Educ.* 26, 605–614. doi: 10.1080/02602930120093922
- Thomas, L., and Jones, R. (2017). *Student Engagement in the Context of Commuter Students*. London: The Student Engagement Partnership (TSEP). Available online at: [http://www.lizthomasassociates.co.uk/projects/2018/Commuter student engagement.pdf](http://www.lizthomasassociates.co.uk/projects/2018/Commuter%20student%20engagement.pdf)
- Thomas, L., and Quinn, J. (2007). *First Generation Entry into Higher Education*. Maidenhead: Open University Press.
- 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. *Oxford Rev. Educ.* 38, 323–342. doi: 10.1080/03054985.2012.689693
- van Heerden, M., Clarence, S., and Bharuthram, S. (2017). What lies beneath: exploring the deeper purposes of feedback on student writing through considering disciplinary knowledge and knowers. *Assess. Eval. High. Educ.* 42, 967–977. doi: 10.1080/02602938.2016.1212985
- van Merriënboer, J. J. G., and Sweller, J. (2005). Cognitive load theory and complex learning: recent developments and future directions. *Educ. Psychol. Rev.* 17, 147–177. doi: 10.1007/s10648-005-3951-0
- Vermunt, J. D., and Verloop, N. (1999). Congruence and friction between learning and teaching. *Learn. Instr.* 9, 257–280. doi: 10.1016/S0959-4752(98)00028-0
- Vogel-Walcutt, J. J., Gebirim, J. B., Bowers, C., Carper, T. M., and Nicholson, D. (2011). Cognitive load theory vs. constructivist approaches: which best leads to efficient, deep learning? *J. Comput. Assist. Learn.* 27, 133–145. doi: 10.1111/j.1365-2729.2010.00381.x
- Vygotsky, L. S. (1978). *Mind in Society: Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
- Waring, M., and Evans, C. (2015). *Understanding Pedagogy: Developing a Critical Approach to Teaching and Learning*. Abingdon: Routledge.
- Winstone, N. E., Nash, R. A., Parker, M., and Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of reciprocity processes. *Educ. Psychol.* 52, 17–37. doi: 10.1080/00461520.2016.1207538

- Witkowsky, P., Mendez, S., Ogunbowo, O., Clayton, G., and Hernandez, N. (2016). Nontraditional student perceptions of collegiate inclusion. *J. Contin. High. Educ.* 64, 30–41. doi: 10.1080/07377363.2016.1130581
- Yucel, R., Bird, F. L., Young, J., and Blanksby, T. (2014). The road to self-assessment: exemplar marking before peer review develops first-year students' capacity to judge the quality of a scientific report. *Assess. Eval. High. Educ.* 39, 971–986. doi: 10.1080/02602938.2014.880400
- Yuriev, E., Naidu, S., Schembri, L. S., and Short, J. L. (2017). Scaffolding the development of problem-solving skills in chemistry: guiding novice students out of dead ends and false starts. *Chem. Educ. Res. Pract.* 18, 486–504. doi: 10.1039/C7RP00009J
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: an overview. *Theory Pract.* 41, 64–70. doi: 10.1207/s15430421tip4102\_2

**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 Balloo, Evans, Hughes, Zhu and Winstone. 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.



# Corrigendum: Transparency Isn't Spoon-Feeding: How a Transformative Approach to the Use of Explicit Assessment Criteria Can Support Student Self-Regulation

Kieran Balloo<sup>1\*</sup>, Carol Evans<sup>2</sup>, Annie Hughes<sup>3</sup>, Xiaotong Zhu<sup>2</sup> and Naomi Winstone<sup>1</sup>

<sup>1</sup> Department of Higher Education, University of Surrey, Guildford, United Kingdom, <sup>2</sup> Southampton Education School, University of Southampton, Southampton, United Kingdom, <sup>3</sup> Learning and Teaching Enhancement Centre, Directorate for Student Achievement, Kingston University, Kingston upon Thames, United Kingdom

**Keywords:** assessment, feedback, criteria, higher education, inclusive curriculum, self-regulation, spoon-feeding, transparency

## A Corrigendum on

## OPEN ACCESS

### Approved by:

Frontiers in Education Editorial Office,  
Frontiers Media SA, Switzerland

### \*Correspondence:

Kieran Balloo  
k.balloo@surrey.ac.uk

### Specialty section:

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

**Received:** 07 September 2018

**Accepted:** 11 September 2018

**Published:** 03 October 2018

### Citation:

Balloo K, Evans C, Hughes A, Zhu X  
and Winstone N (2018) Corrigendum:  
Transparency Isn't Spoon-Feeding:  
How a Transformative Approach to  
the Use of Explicit Assessment Criteria  
Can Support Student Self-Regulation.  
*Front. Educ.* 3:85.  
doi: 10.3389/feduc.2018.00085

## Transparency Isn't Spoon-Feeding: How a Transformative Approach to the Use of Explicit Assessment Criteria Can Support Student Self-Regulation

by Balloo, K., Evans, C., Hughes, A., Zhu, X., and Winstone, N. (2018). *Front. Educ.* 3:69.  
doi: 10.3389/feduc.2018.00069

In the original article, we use the phrase “criteria compliance” without citing Torrance (2007). Citations have now been added to the relevant sections and the updated paragraphs appear below.

## ABSTRACT

If little care is taken when establishing clear assessment requirements, there is the potential for spoon-feeding. However, in this conceptual article we argue that transparency in assessment is essential to providing equality of opportunity and promoting students' self-regulatory capacity. We begin by showing how a research-informed inclusive pedagogy, the *EAT Framework*, can be used to improve assessment practices to ensure that the purposes, processes, and requirements of assessment are clear and explicit to students. The *EAT Framework* foregrounds how students' and teachers' conceptions of learning (i.e., whether one has a transactional or transformative conception of learning within a specific context) impact assessment practices. In this article, we highlight the importance of being explicit in promoting access to learning, and in referencing the *EAT Framework*, the importance of developing transformative rather than transactional approaches to being explicit. Firstly, we discuss how transparency in the assessment process could lead to “criteria compliance” (Torrance, 2007, p. 282) and learner instrumentalism if a transactional approach to transparency, involving high external regulation, is used. Importantly, we highlight how explicit assessment criteria can hinder learner autonomy if paired with an overreliance on criteria-focused ‘coaching’ from teachers. We then address how ‘being explicit with

assessment' does not constitute spoon-feeding when used to promote understanding of assessment practices, and the application of deeper approaches to learning as an integral component of an inclusive learning environment. We then provide evidence on how explicit assessment criteria allow students to self-assess as part of self-regulation, noting that explicit criteria may be more effective when drawing on a transformative approach to transparency, which acknowledges the importance of transparent and mutual student-teacher communications about assessment requirements. We conclude by providing recommendations to teachers and students about how explicit assessment criteria can be used to improve students' learning. Through an emphasis on transparency of process, clarity of roles, and explication of what constitutes quality within a specific discipline, underpinned by a transformative approach, students and teachers should be better equipped to self-manage their own learning and teaching.

## INTRODUCTION, Paragraph 2

If little care is taken when establishing clear assessment requirements, there is the potential for "spoon-feeding," yet the move toward transparency in assessment in higher education has largely been positively received (Carless, 2015), since explicit requirements are likely to facilitate fairness in marking practices by enhancing markers' abilities to be consistent in making accurate judgments of student work (Broadbent et al., 2018) and communicating reasons for a particular judgment (Sadler, 2005). Explicit assessment criteria can support students to consider what they are aiming for and how this can be achieved from the perspective of a marker (Nicol and MacFarlane-Dick, 2006), so their learning outcomes move beyond a purely cognitive product, to the development of metacognition (Frederiksen and

Collins, 1989; Shephard, 2000; Swaffield, 2011) and assessment literacy (Price et al., 2012). In this article, we present a conceptual analysis of the value of explicit assessment criteria; we highlight the potential risk of spoon-feeding in promoting "criteria compliance" (Torrance, 2007, p. 282), and then we present approaches demonstrating that a careful use of transparency through explicit assessment criteria is crucial to promoting equality of opportunity and students' self-regulation.

## AT RISK OF SPOON-FEEDING? A TRANSACTIONAL APPROACH TO TRANSPARENCY IN ASSESSMENT PRACTICES, Paragraph 5

In the context of assessment, spoon-feeding may involve explicitly telling students *what* they need to do for an assignment, and *how* to meet the assessment criteria, without leaving it up to them to ascertain this for themselves. Addressing task criteria in the absence of understanding the domain being assessed has been termed by Torrance (2007) as "criteria compliance" (p. 282). Some students may use explicit criteria to focus on exactly what needs to be done to reach a desired level of achievement, rather than actually learning material fully (Panadero and Jonsson, 2013). Students' and teachers' conceptions of learning play a role in this; if teachers simply supply assessment requirements to students in a transactional manner, so they can passively "check boxes," it is unlikely that students will engage with the criteria in a way that will develop their learning and self-regulation.

The authors apologize for this error and state that this does not change the conceptual analysis or conclusions presented in the article in any way.

The original article has been updated.

## REFERENCES

- Broadbent, J., Panadero, E., and Boud, D. (2018). Implementing summative assessment with a formative flavour: a case study in a large class. *Assess. Eval. High. Educ.* 43, 307–322. doi: 10.1080/02602938.2017.1343455
- Carless, D. (2015). Exploring learning-oriented assessment processes. *High. Educ.* 69, 963–976. doi: 10.1007/s10734-014-9816-z
- Frederiksen, J. R., and Collins, A. (1989). A systems approach to educational testing. *Educ. Res.* 18, 27–32. doi: 10.3102/0013189X018009027
- Nicol, D., and MacFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Stud. High. Educ.* 31, 199–218. doi: 10.1080/03075070600572090
- Panadero, E., and Jonsson, 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
- Price, M., Rust, C., O'Donovan, B., Handley, K., and Bryant, R. (2012). *Assessment Literacy: The Foundation for Improving Student Learning*. Oxford: Oxford Centre for Staff and Learning Development.
- Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assess. Eval. High. Educ.* 30, 175–194. doi: 10.1080/0260293042000264262
- Shephard, L. A. (2000). The role of assessment in a learning culture. *Educ. Res.* 29, 4–14. doi: 10.3102/0013189X029007004
- Swaffield, S. (2011). Getting to the heart of authentic assessment for learning. *Assess. Educ. Princ. Policy Pract.* 18, 433–449. doi: 10.1080/0969594X.2011.582838
- 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

**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 Balloo, Evans, Hughes, Zhu and Winstone. 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.





# From “Seeing Through” to “Seeing With”: Assessment Criteria and the Myths of Transparency

Margaret Bearman\* and Rola Ajjawi

Centre for Research in Assessment and Digital Learning, Deakin University, Geelong, VIC, Australia

The notion of “transparency” has been extensively critiqued with respect to higher education. These critiques have serious implications for how educators may think about, develop, and work with assessment criteria. This conceptual paper draws from constructivist and post-structural critiques of transparency to challenge two myths associated with assessment criteria: (1) transparency is achievable and (2) transparency is neutral. Transparency is interrogated as a social and political notion; assessment criteria are positioned as never completely transparent texts which fulfill various agendas. Some of these agendas support learning but this is not inevitable. This conceptual paper prompts educators and administrators to be mindful about how they think about, use, and develop assessment criteria, in order to avoid taken-for-granted practices, which may not benefit student learning.

## OPEN ACCESS

### Edited by:

Anders Jönsson,  
Kristianstad University, Sweden

### Reviewed by:

Sue Bloxham,  
University of Cumbria,  
United Kingdom  
Gavin T. L. Brown,  
University of Auckland, New Zealand

### \*Correspondence:

Margaret Bearman  
margaret.bearman@deakin.edu.au

### Specialty section:

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

Received: 01 May 2018

Accepted: 18 October 2018

Published: 05 November 2018

### Citation:

Bearman M and Ajjawi R (2018) From  
“Seeing Through” to “Seeing With”:  
Assessment Criteria and the Myths of  
Transparency. *Front. Educ.* 3:96.  
doi: 10.3389/feduc.2018.00096

**Keywords:** assessment criteria, Transparency, higher education, standards, rubrics

## INTRODUCTION

In higher education, it is generally considered desirable for assessment criteria to be “transparent” (Jackel et al., 2017). In this sense, transparency means that educators are explicit about their expectations for assessment and students therefore can see what it is they need to achieve. For many, a significant reason for providing transparent criteria is to help students learn. Jonsson (2014, p. 840) summarizes this approach as: “Student awareness of the purpose of the assessment and assessment criteria is often referred to as *transparency* ... in order to educate and improve [a] student’s performance, all tasks, criteria and standards must be *transparent* to both students and teachers.” [italics ours] However, transparency as a concept may be more than it seems. The complexities and nuances of the transparency agenda have been explored and critiqued with respect to higher education in general (Strathern, 2000; Brancalone and O’Brien, 2011; Jankowski and Provezis, 2014) and assessment in specific (Orr, 2007), primarily through a post-structural lens. To the best of our knowledge, this previous work has not directly concerned the transparency of assessment criteria. In landscapes where the use of rubrics have become taken-for-granted, it is worth interrogating more closely some of the underpinning assumptions around transparency of assessment criteria.

This paper seeks to overturn myths associated with transparency of assessment criteria. We challenge the notion that transparent assessment criteria are (a) possible and (b) an unqualified good. While we draw from published critiques of transparency, we are not calling for wholesale abandonment of explicating criteria in text; we acknowledge that the notion of transparent assessment criteria serves valuable purposes in making teachers accountable and in providing direction for students. Rather, we suggest that the way transparency is enacted in assessment criteria

in the daily practice of university teaching and learning, may not take account of its limitations. To make this argument, we outline the general landscape of written assessment criteria in higher education. We then problematise the notion of transparent assessment criteria, with particular attention to these two myths. Our arguments are illustrated with a critical examination of a bioethics rubric. We do not choose this example to highlight flaws with a particular rubric design, but to illuminate how the notion of transparency might lead to poor use of rubrics. Finally, we explore implications by offering some considerations for educators, managers and quality improvement staff when developing or working with rubrics.

## WRITTEN ASSESSMENT CRITERIA IN HIGHER EDUCATION

In higher education, transparency of assessment criteria is part of a larger movement from assessment being “secret teachers’ business” to something that is made public to students and the wider community (Boud, 2014). In particular, tertiary institutions in countries such as Australia and the United Kingdom have moved to the explicit articulation of course and unit learning outcomes. Assessment and associated criteria present a means for assuring that the students can meet these learning outcomes. This is part of a significant change in assessment practice, whereby students are graded against a standard rather than against each other (Sadler, 2009).

University assessors judge student work against a series of criteria (Sadler, 2009), which reference academic standards. In order for assessment criteria (or standards) to be “transparent,” they are recorded, generally in writing, and shared between students, educators and administrators. Increasingly, explicit written assessment criteria take the form of rubrics; these are a pervasive presence in the higher education literature (Dawson, 2015). This literature suggests that: students like the provision of rubrics (Reddy and Andrade, 2010); students consider them helpful (Reddy and Andrade, 2010); and that rubrics may improve learning (Panadero and Jonsson, 2013). From a student perspective, teachers sharing these types of written expressions of assessment criteria are the primary means of coming to know the standards for the course or unit. Reading the rubric may be the only time a student will engage or think about the quality of what they are trying to achieve.

In summary, assessment criteria provide judgement points for an assessment task, drawing from academic standards. Through written form such as rubrics, educators seek to make assessment criteria “known” to students. This is also intended to have educative effects on the students and build their awareness of the standard. The written assessment criteria are therefore the focus of this paper, as they are both a ubiquitous part of practice and the means whereby educators seek to achieve “transparency.”

## TRANSPARENCY IS TAKEN-FOR-GRANTED

The taken-for-granted benefits of transparent standards and criteria are a “normalized discourse” in higher education

(Orr, 2007, p. 646). For example, a 2017 literature review of higher education assessment notes: “... when [standards] are clearly articulated, and when students engage with them, performance standards help improve *transparency of assessment* and student learning...” [italics ours] (Jackel et al., 2017, p. 18). Likewise, (Rodríguez-Gómez and Ibarra-Sáiz, 2015, p. 4) describe transparency as a foundational principle for assessment, noting: “Assessment is carried out against a set of *transparent rules, standards and criteria* which guide students to achieve the required learning outcomes...” [italics ours]. It can be seen from this phrasing that transparency is regarded as a general good. Indeed, transparency is paired with learning as a desirable outcome. And, it could be argued, why not? As mentioned, there is evidence that clear criteria encapsulated in rubrics help both educators’ communication of the standards and students’ learning (Reddy and Andrade, 2010; Panadero and Jonsson, 2013; Jonsson, 2014). So then, why should we question it? Does it not benefit ourselves and our students to clearly articulate what it is they are supposed to do?

We suggest that by thinking more deeply about transparency, we can improve the way we use assessment criteria in our teaching. There is significant post-structural critique regarding the discourse of transparency, indeed transparency has already been problematized with respect to assessment and higher education (Orr, 2007; Jankowski and Provezis, 2014). Likewise there has been extensive acknowledgements of the inherent challenges of being explicit (O’Donovan et al., 2004; Sadler, 2007; Torrance, 2007). However, experienced educators are not necessarily aware of this literature when they work with standards or criteria (Hudson et al., 2017). We think it is therefore necessary to interrogate the taken-for-granted nature of transparency with specific reference to assessment criteria.

We challenge two myths about transparency in order to help express assessment criteria more productively. The first myth is that transparency is achievable and the second is that transparency is neutral.

## MYTH 1: TRANSPARENCY IS ACHIEVABLE

Possibly the most pervasive assumption about any form of transparency is that it makes everything visible, like shining a light into a dark room. This may mean that when academics invoke transparency with respect to assessment criteria, they sometimes assume that there are objective standards, which can be precisely and accurately described. Academic standards, however, have been acknowledged to be social constructions, which have a “necessarily elusive and dynamic nature ... continuously co-constructed by academic communities and ferociously difficult to explain to a lay audience.” (Bloxham and Boyd, 2012, p. 617). Already this notion of dynamic and tacit standards necessarily challenges the notion of making everything visible. While it could be argued that written assessment criteria is a means of making our social truths explicit, this seems to somewhat miss the point. We think there are other, more complex forces at work. We offer three arguments that suggest that rubrics and similar can never make everything visible.

### 1) *There is knowledge that cannot be expressed*

One of the most common challenges in writing assessment criteria is capturing holistic tacit knowledge; and many argue that this knowledge is impossible to capture explicitly (O'Donovan et al., 2004; Orr, 2007; Sadler, 2009; Bloxham and Boyd, 2012; Bloxham et al., 2016; Hudson et al., 2017). Any expression of standards and criteria necessarily simplifies and clarifies the complex nature of work, in order to communicate it. That is, by capturing knowledge in words, we lose some sense of it. O'Donovan et al. (2004) describe how they once believed that: "making assessment criteria and standards transparent ... could be achieved fairly simply through the development and application of explicit school-wide assessment criteria and grade descriptors." (O'Donovan et al., 2004, p. 327) However, they came to learn that tacit knowledge was impossible to pin down, despite considerable effort.

### 2) *Transparent criteria are in the eye of the beholder*

We suggest that if academic standards are socially constructed and based on tacit, dynamic knowledge, then how these standards are perceived and how the knowledge is understood, depends on an individual's social history and standing. This applies equally to assessment criteria. As Jankowski and Provezis (2014, p. 481) note: "...the request for assessment information to be transparent is challenging because employers, students, institutions and policy makers have different understandings of ... what it means to be transparent. In other words, what may be transparent for one group may not be for another." This comes to a matter of interpretation whereby we bring our frames of reference to make sense of criteria (Tummons, 2014).

How students make sense of criteria may inform their perspective as to what constitutes "transparent" criteria. A study of students' perspectives of assessment criteria suggests students have divergent ways of engaging with criteria (Bell et al., 2013). On the one hand, there were those students who wished to use the rubric as a recipe; and on the other, there were those who embraced a more complex idea of standards, closer to an educator's perspective. These two groups spoke about how they interacted with the written assessment criteria in very different ways. We suggest that for some students, the notion of transparency related to the assignment, and for other students, transparency related to the underlying standards.

Moreover, the written expression of the criteria can be interpreted in diverse ways, depending on how much the student already knows. Often students, as novices in the field, may not be able to make sense of the language used in rubrics or similar as they do not have the necessary repertoires of understanding (which may be developed during their studies). In other words, students' ability to "see through" to the assessment criteria, depends on their *a priori* knowledge; what is transparent for an expert may be opaque to the student. How much a rubric can prompt understanding and learning is therefore dependent on the student as much as the transparency of the written criteria.

### 3) *Making some criteria transparent makes other criteria opaque*

Strathern's (2000) critique of transparency of audit in higher education asks the question: "what does visibility conceal?" This

is one of the key challenges to transparency: you cannot make a choice about what you say, without making a choice about what you do not say. The text of any written assessment criteria suggests that particular forms of knowledge are particularly important: students should be paying attention to *this*. In doing so, without mentioning it, they direct students' attention away from *that*. So the whole notion of transparency must inevitably be based on highlighting some things and *obscuring others*. To use post-structural language, making assessment criteria transparent both legitimizes and delegitimizes particular forms of knowledge. We suggest this indicates that transparency is not neutral—and this is exactly the point that is explored within our next myth.

## MYTH 2: TRANSPARENCY IS NEUTRAL

Orr (2007) argues that the discourse of assessment in higher education is mostly rooted in positivism, with its emphasis on attainment of measurable standards that are constant over time. "Transparency" stems from this discourse, which is underpinned by the notion that standards are knowable, expressible and measurable; we have previously described flaws in this perspective. Alternative discourses position assessment as a socio-political act (Orr, 2007; Raaper, 2016). The transparency movement can therefore be seen as part of a political system. This is not in itself "bad;" after all, the need for transparency is what prevents our assessments from being "secret" (Boud, 2014) and assessors from abusing their authority (Raaper, 2016) as well as offering students a more equal footing (Ajjawi and Bearman, 2018). However, this is also a system where teachers feel pressurized and constrained by assessment (Raaper, 2016) and students "game the system" (Norton, 2004).

We suggest that it is worth being cautious about seeing transparency as a benefit in and of itself. Neoliberal critiques suggest that "transparency" can also be considered a form of scrutiny (Strathern, 2000). As mentioned above, we do not see "scrutiny" as an evil or even a necessary evil; scrutiny also ensures that assessments are not deliberately obscured and abused. However, we propose that various groups can use the sharing of written assessment criteria to fulfill diverse agendas and this use of transparent written criteria can be seen as a form of control. Hence it is worth asking: which agendas do the written criteria serve? We provide three ways of viewing transparent assessment criteria, which challenge their apparent neutrality.

### 1) *Transparent assessment criteria as governance*

For some, the very notion of transparency can be seen as contributing to a system that seeks to commodify and control education (Brancaleone and O'Brien, 2011). From this view, transparency enacted through written assessment criteria permit institutions to enforce governance. By ensuring assessment criteria are visible, institutions have a means to control teachers and teaching (Jankowski and Provezis, 2014). We suggest that this may lead to positive learning outcomes, as transparent criteria allow courses, faculties and institutions to ensure that standards are met. However, it may also lead to problems with teaching and therefore, learning. For example, in an attempt to secure comparability and equity, the same rubric may be

used across different tasks in a course or unit or even an entire discipline. Attempts to provide these comparable criteria can sometimes produce deep frustration from assessors, lecturers and students if the generic approaches do not capture the nuance of the task at hand. Hudson et al. (2017) describes the need for “conceptual acrobatics,” where teachers’ professional judgements about what to address at a particular time with particular students is in conflict with “transparent” and therefore pre-set learning outcomes. This illustrates how, on occasion, a desire for comparable and transparent standards might lead to the de-professionalization of teaching.

## 2) *Transparent assessment criteria as a means of student control (and control of students)*

Written assessment criteria allow the “secret assessment business” to become shared and public. In this way, making assessment criteria transparent cedes control of assessment to students. This is generally a good thing: through reading and working with rubrics or similar, students become able to form their own perspectives on why and how their work meets the standard. Ideally, we embrace this and allow students to co-construct rubrics with teachers. In this process, students and teachers work together to express the criteria by which work will be judged (Fraile et al., 2017). We would regard this as an example of where transparent assessment criteria is beneficial to learning.

The students themselves can use the written criteria to control their own experiences. Many students seek to use the written criteria to pass the assessment rather than learn (Norton, 2004; Bell et al., 2013). Academics are all familiar with students who come, checklist in hand, saying “why did I get this mark?” Sometimes this is valuable, and sometimes students are arguing the letter rather than the spirit of the text. This is another consequence of seeking transparency, which is not necessarily aligned with learning, but which illustrates how visible assessment criteria give control to students.

However, seeking transparency may also prompt educators to tightly specify levels of achievement; in some instances, this may be a misguided attempt to control students’ outputs in a way that helps them meet the criteria, but not necessarily learn (Torrance, 2007). The danger here is that seeking transparency is conflated with reductiveness: “fine-grained prescription, atomised assessment, the accumulation of little ‘credits’ like grains of sand, and intensive coaching toward short-term objectives, are a long call from the production of truly integrated knowledge and skill.” (Sadler, 2007, p. 392) We do not suggest reductive expressions of criteria are *necessarily* the result of making assessment “transparent,” but that they remain a *possibility* (and thus, frequently, a reality).

## 3) *Transparency controls how students see knowledge*

As mentioned, the standards to which our criteria refer, can be considered social constructs. They shift and change over time. By invoking transparency as simply “seeing through,” we may inadvertently create the notion that criteria are fixed, durable and objective (Ajjawi and Bearman, 2018). However, this is not the case: we have already described how some forms of knowledge are permitted whilst others are constrained. This is not a good

or bad thing in itself, although it can lead to both desirable and undesirable outcomes. For example, in the 1800s, the gold standard for being a lawyer might have included being a man from a certain class. No matter how clear or transparent the criteria, these standards would not be defensible in our current day and age. Equally, our 2018 criteria may reflect unconscious social attitudes of our era. On the other hand, criteria can deliberately try to drive social change. For example, medical curricula that have holistic criteria around professionalism may deliberately be seeking to change the discourse around what it means to be “a good doctor.”

Our discourses of transparency may also affect how students see knowledge. Saying that our assessment criteria are “transparent” reveals how we understand knowledge itself, or our epistemic beliefs, and therefore power. From an educational perspective, what we want to avoid, is giving our students the sense that knowledge is fixed and stable. Higher education aspires to develop students’ personal epistemologies (Hofer, 2001), so that students can come to understand multiple views and the dynamic nature of knowledge. How we talk about our assessment criteria and enact transparency may influence how our students come to understand what knowledge means.

## WHAT THE MYTHS OF “TRANSPARENCY” MEAN ON THE GROUND: THE BIOETHICS RUBRIC EXAMPLE

Taken together, we have mounted a case that transparent assessment criteria are not true reflections of some kind of objective truth but texts that are necessarily open to interpretation and which allow control and scrutiny. We use an extract from a bioethics rubric (Hack et al., 2014; Hack, 2015) to explore how these critiques of transparency collide with the “taken-for-granted” nature of transparent assessment criteria. That is, we look at the tensions between a socio-political frame and assessment criteria as explicit guidance from teachers to learners about the assessment requirements.

### The Sample Rubric

For this purpose, the sample rubric is not intended to illustrate outstanding or particularly poor practice. Rather, it represents an “on the ground” illustration of a thoughtfully written assessment criterion—typical of many rubrics. The whole (originally co-constructed) rubric is available through a Creative Commons license (Hack, 2015). We present one row—“ethics”—in **Table 1**. It relates to the following assignment brief: “Prepare a critical examination of the key issues on a current topic in medical or health science which raises ethical issues. You should draw extensively on the literature to present: an introduction to the technology or science that underpins the issue, the key ethical aspects with reference to ethical theory, the implications for policy decisions, practice and/or regulatory frameworks.” (Hack et al., 2014) The following analysis illustrates how these different notions of transparency are enacted within this small piece of text, and from this, we draw out implications for learning and



**TABLE 1** | Sample row from bioethics rubric reproduced from Hack (2015).

|               | Fail   | Pass   | Merit   | Distinction  |
|---------------|--|--|---|--|
| <b>Ethics</b> | <b>0–9 points</b><br>Missing key stakeholders, or issues not explored or incorrect. Some evaluation and synthesis of issues and material. Basic knowledge and understanding of key principles justice, autonomy, and well-being. | <b>10–11 points</b><br>Identifies stakeholders and important issues, demonstrates some competence in analyzing and interpreting an ethical issue. Communicates understanding of the ethical principles: justice, autonomy, and well-being. | <b>12–13 points</b><br>Identifies stakeholders and important issues, demonstrates competence in analyzing and interpreting an ethical issue. Communicates understanding of the ethical principles: justice, autonomy, and well-being. | <b>14–20 points</b><br>Identifies stakeholders and issues, demonstrates a very high level of competence in analyzing and interpreting a complex ethical issue, potentially with unforeseen circumstances. Communicates understanding of the ethical principles: justice, autonomy, and well-being. |

teaching. We have deliberately chosen a row pertaining to a complex aspect of the task.

## Illustration of the Arguments

In the following sections, we draw out and illustrate how the text of this rubric supports our earlier contentions. Italics indicate the key arguments as discussed in previous sections.

The assumption underpinning the notion of transparent assessment criteria is that, by educators expressing what students need to achieve, students will come to know the standards behind the criteria and hence be able to improve their performance on the assessment task (i.e., “see through” the criteria to the standards). At broad brushstroke, this co-constructed written criterion (**Table 1**) may meet this aim. However, there are several different sorts of complex knowledge that are necessarily simplified and represented in this text. There are references to “ethical issues” and “key principles of justice, autonomy and well-being,” which pertain to the content knowledge of the course. There are verbs that articulate the type of generic intellectual acts that students are supposed to achieve, such as “synthesise” and “evaluate.” There are also descriptors of the expected standard: “basic,” “some competence” and “very high level of competence.” All these different elements come together in the statement “demonstrates some competence in analyzing and interpreting an ethical issue.” This statement contains a great deal of tacit knowledge. In other words, while the rubric comes to some sense of what is intended, it does not (and cannot) explicate it entirely as it contains *knowledge that is not documentable*. This has been noted beforehand (O’Donovan et al., 2004; Sadler, 2007) but it is worth repeating as the literature shows that educators can make the assumption (O’Donovan et al., 2004; Hudson et al., 2017) that: (a) if it is in the rubric, then assessors and students should know what is expected and b) if they expend enough energy on explicating it, then there will be clarity about the standards. However, these are part of the myth of *transparency being achievable*. Others have suggested useful means to support students to come to know standards beyond telling, such as the use of exemplars (Carless et al., 2018). Likewise, the act of co-construction comes a long way toward addressing “knowing” that isn’t the same as “making transparent.”

Exemplars, co-construction and other pedagogical supports can also help with challenges resulting from *transparency being in the eye of the beholder*. Let us consider a statement from

the distinction column in **Table 1**: “...demonstrates a very high level of competence in analyzing and interpreting a complex ethical issue, potentially with unforeseen circumstances.” So, for the educator, the rubric expresses a general (but not complete) view of what they are expecting. For one student, with a good repertoire of ethics and a particular set of epistemic beliefs, this rubric may fulfill some educative purpose. That is, through reading the rubric, they may learn that part of ethical issues is trying to forecast “unforeseen circumstances.” On the other hand, a student who is more novice may struggle to grasp what an “unforeseen circumstance” might even be. Similarly, “a high level of competence” may mean entirely different things to different people depending on their own competence in the area (Kruger and Dunning, 1999). The most significant implication of assessment criteria being read differently by different people, is that written criteria are best understood by those who already have the knowledge to fulfill the requirements. In other words, those who most need to learn what the standards are, are likely more mystified by assessment criteria than those who already have a grasp of the course content. This again returns to the point that providing the rubric is insufficient: the concepts within it must be supported by the rest of the curriculum as well as assessment artifacts such as exemplars. In our particular example, the rubric is co-constructed: in this way, students and staff jointly contribute to the tacit understandings underpinning the rubric and so come to a shared view of the expectations set by the assessment criteria. However, this only mitigates the challenge for those immediately involved in the construction, it does not remove it.

We have suggested that assessment criteria always operate in a broader social landscape. For example, from a content perspective, these assessment criteria legitimate a certain view of ethics. The decision to focus on “justice, autonomy, and well-being” and not other ethical frameworks may be an appropriate educator choice against the curricula, but what is being excluded is not captured within the written criteria and hence is made invisible. Similarly, the assessment—and the criteria—emphasize an academic discussion, not a personal reflection or some other mode of expression. Again, this legitimizes the academic form and delegitimizes the alternatives. We do not suggest that there is anything wrong with this—indeed it is inevitable—but to make the point that making *some criteria transparent*, makes other criteria *opaque*. Once again this comes to the point

that transparency is not achievable, but it also underlines our argument that *transparency is not neutral*.

We now turn to a more political frame. Drawing again from our illustration in **Table 1**, we seek to explore *transparency as governance*. There are many ways in which assessment criteria like these permit control and scrutiny of teacher activity. For example, this row in the rubric could be provided in an audit as evidence of meeting requirements to teach bioethics. As another example, institutional policy might dictate that changes to units of study, often including rubrics, must be submitted to scrutiny by committee almost a year in advance to ensure they represent the university sanctioned standards and thereby limiting teacher agility and control. Rubrics also work as coordinating agents across classes, students and teachers—to ensure consistency. While this consistency is at best fleeting (Tummons, 2014), the rubric may limit the ability of teachers to modify assessment criteria to incorporate discussions with a particular cohort. This is not to say that all control and scrutiny leads to negative consequence. Moderation relies on these types of rubrics to coordinate and control grading processes. If this rubric was used to promote a discussion of what educators regarded as meeting the “distinction” criteria across the teaching team, then we would see this as a valuable form of governance, which ultimately leads to better teaching and learning. Note that there is an overlay between *transparency as governance* and the recognition that *some knowledge is not documentable*. If institutions believe that transparency is achievable, then this may create tensions in how the assessment criteria are used for governance.

While transparent assessment criteria may provide a means for institutional “scrutiny” and “control,” they also allow *student control and control of students*, depending on the broader institutional context. In our example, as a co-constructed rubric, students necessarily have some control over these criteria; it promotes their ownership and investment as well as their understanding. Key terms, such as “stakeholders” and “ethical issues” are likely to have been discussed and possibly argued for. As mentioned, they also can use these criteria to query grades (“I have ‘demonstrated some competence in analyzing and interpreting an ethical issue’ why aren’t I passing?”).

However, assessment criteria equally control students. As is illustrated by our one row from the rubric, they highlight what the student should privilege. What is now deemed important in the assignment are: discussion of stakeholders, analysis and interpretation of ethical issues; and communication of “understanding” ethical principles. Students will now fulfill this. This is not necessarily problematic—indeed it is arguably the point of education—but it may become so if the assessment criteria are prescribed so tightly, they become a recipe book. For example, in our illustrative bioethics assignment, it would be entirely counterproductive to replace this rubric with a series of very tight criteria regarding analysis of a complex ethical issue; the point of learning ethical thinking is to manage complex nuance. We would suggest that, outside of particular focussed skills such as resuscitation or pipetting, analytical (atomised) written criteria should allow students to make their own meanings, and develop their own judgements about what constitutes quality. Another means whereby a rubric may

control students, is if it is provided to students with minimal explanation, then feedback returned to students again with minimal explanation, simply using the text in the rubric. In these circumstances, students are primed to explicitly follow the rubric thus limiting creativity and exploration of other forms of ethical practice. Here the use of the rubric is inadvertently *controlling how students see knowledge*.

In our view, our illustrative rubric mostly supports a dynamic view of knowledge. The process of co-construction reveals the dynamic nature of criteria, which are devised and developed through consensus and discussion. Moreover, the verbs “communicates,” “analyses” and “interprets” positions knowledge as constructed rather than “hard facts.” On the other hand, the language of this criterion does present some absolutes. For instance, it subtly but distinctly privileges “justice, autonomy and well-being” as “the” ethical principles; whether they are or not is not the question at hand. What we are underlining is that the text suggests that a framework is fixed and durable. In short, written criteria contain subtle (and not so subtle) messages about the nature of content knowledge in specific and epistemology in general, and that as educators we should try and take account of this.

## Implications of the Illustrative Example

This examination of a small sample of text shows that a single row in the rubric can fulfill one part of the transparency agenda—by ensuring that educators’ broad expectations are communicated to students and students come to some understanding of the underlying standards. This can be done without “complete” transparency; the text is not completely explicit, and nor does it need to be. However, from a socio-political perspective, the rubric is doing so much more. It allows institutions, educators and students to control each other. This is not necessarily a bad thing, what we suggest is critical is that educators (and ultimately students) be more aware of how objects such as rubrics operate within the educative space. The overall implication is that it may be more useful for educators to think about how the criteria will be used and by whom, when they are considering how the written criteria reflect the tacit standard. In other words, how will students (and teachers) “see with” the criteria, rather than “see through” them.

## FROM “SEEING THROUGH” TO “SEEING WITH”

We have mounted a case against a simplistic understanding of assessment criteria that sees transparency as an unqualified good. We suggest that discourses and processes of transparency are not a matter of expressing with greater clarity. Instead, written assessment criteria form part of a much larger social and political landscape. “Making transparent” is neither neutral, nor in fact, possible. So what does all this mean for assessment practice?

There are considerable implications for how we develop and use assessment criteria. The first inference is that educators can reflexively consider their own views of assessment criteria and

transparency, and how these are reflected (or not) in what they do. To give an example of this type of reflexive thinking, we will describe what we ourselves as educators see as desirable. We would like our assessment criteria to allow our students to make their own meanings about work in relation to holistic dynamic standards (Ajjawi and Bearman, 2018). We design processes to incorporate the written assessment criteria into teaching, building on all the tasks (and associated feedback opportunities) that we have provided during teaching, such as through formative self and peer assessment. At the same time, we want written assessment criteria to provide ourselves, our fellow markers and our students, a shared sense of what we think is good work. We also want to open ourselves to the opportunity for our students to teach us about the standards, not just the other way around. To this end, we want ourselves and our students to “see with” criteria, not “see through” them.

Ways to achieve our aspirational use of assessment criteria are already described within the literature (e.g., Norton, 2004; O'Donovan et al., 2004). Sadler (2009) writes: “Bringing students into a knowledge of standards requires considerably more than sending them one-way messages through rubrics, written feedback or other forms of telling. It requires use of the same tools as those employed for setting, conveying and sharing standards among teachers: exemplars, explanations, conversation and tacit knowing.” (Sadler, 2009, p. 822) While these dialogs are how expert educators help their students make meaning of standards, these clearly are not the same as “seeing through” to the criteria. Designing activities for students to “see with” criteria, (i.e., the pedagogical activities that support a rubric), goes a long way toward the development of a shared understanding of the standards for an assessment.

We also recognize that the pragmatics of assessment design may mean that control over assessment tasks and associated criteria may not be possible (Bearman et al., 2017). However, even in the most constrained of circumstances, educators (and students) can discuss their criteria with both students and colleagues. With respect to students, the explanation of the criteria can be important for “transparency” (Jonsson, 2014), and may provide an opportunity to enhance or adjust already established tasks and rubrics. With respect to colleagues, it can highlight the role the criteria take in shaping and controlling student and educator behavior. It is easy to fall into assessment practices that are the “way things are done around here” (Bearman et al., 2017). Critically examining the words for taken-for-granted assumptions may be one way to ensure we do not automatically continue unproductive practices.

To this end, we suggest three framing questions by which educators can approach assessment criteria. These allow them to align with the transparency discourses but also allow them to think more deeply about what “being transparent” might entail. Questions to guide educators are:

- 1) What is my agenda with the text of this criteria and what other agendas might these written criteria serve?
- 2) What might be learnt from these written criteria about the nature of knowledge?

- 3) By what means (e.g., activities, assessment designs, dialogs) will students be given the opportunity to “see with” written criteria?

These questions do not seek to make assessment criteria “transparent” but provide a process to challenge assumptions and promote student benefit. Through this, educators may find a means to adapt notions of “transparency” to their own contexts.

There are also more radical implications of this work. If transparency is unattainable, then reflexivity, criticality and co-construction are necessarily implied when shifting to “seeing with” rather than “seeing through.” Quality assurance processes could seek to explore the educational processes built in and around the assessment criteria, rubrics and standards. From this perspective, written assessment criteria should be accompanied by co-constructive processes that involve discussion of quality, the use of exemplars, and nested formative tasks with dialogic feedback processes that help develop shared understandings of the standards. These processes become the markers of a quality learning and teaching design rather than the presence or otherwise of a rubric (or similar). Similarly, quality assurance processes could favor dialog with staff about the intended and unintended consequences of how such processes play out on the ground—therefore informing future design in a collaborative way. Further, from a political perspective, managers need to attune to the ways in which rubrics can act as disciplinary tools—for example when certain forms are mandated without discussion and where the form the rubric takes to honor consistency trumps professional judgement on the ground. Reflexivity by managers is needed to acknowledge the level of control as well as criticality and humility to question whether such control does in fact lead to better educational processes or merely reduces perceived variability. These implications have real consequences for institutions including an imperative for staff time to be allocated to designing interactive forms of education rather than transmissive ones that aim to do more with less.

## CONCLUSIONS

This paper has explored the nature of transparency with respect to assessment criteria. We have drawn from others' critiques of transparency and assessment criteria to present a case for educators to think about transparency differently. We suggest that transparency is neither *bad* nor *good*, but is a socio-political construction. Therefore, assessment criteria can never be truly transparent, nor would we want them to be. Instead, we may want to ask ourselves: What agendas can assessment criteria serve? How do they direct the students and ourselves? What do they hide?

## AUTHOR CONTRIBUTIONS

MB and RA both contributed to the ideas underpinning this article. MB wrote the primary draft. RA contributed additional text and both revised for important intellectual content.

# REFERENCES

- Ajjawi, R., and Bearman, M. (2018). "Problematising standards: representation or performance?," in *Developing Evaluative Judgement in Higher Education: Assessment for Knowing and Producing Quality Work*, eds D. Boud, R. Ajjawi, P. Dawson, and J. Tai (Abingdon: Routledge), 41–50.
- Bearman, M., Dawson, P., Bennett, S., Hall, M., Molloy, E., Boud, D., et al. (2017). How university teachers design assessments: a cross-disciplinary study. *High. Educ.* 74, 49–64. doi: 10.1007/s10734-016-0027-7
- Bell, A., Mladenovic, R., and Price, M. (2013). Students' perceptions of the usefulness of marking guides, grade descriptors and annotated exemplars. *Assess. Eval. High. Educ.* 38, 769–788. doi: 10.1080/02602938.2012.714738
- Bloxham, S., and Boyd, P. (2012). Accountability in grading student work: securing academic standards in a twenty-first century accountability context. *Br. Educ. Res. J.* 38, 615–634. doi: 10.1080/01411926.2011.569007
- Bloxham, S., den-Outer, B., Hudson, J., and Price, M. (2016). Let's stop the pretence of consistent marking: exploring the multiple limitations of assessment criteria. *Assess. Eval. High. Educ.* 41, 466–481. doi: 10.1080/02602938.2015.1024607
- Boud, D. (2014). Shifting views of assessment: from secret teachers' business to sustaining learning," in *Advances and Innovations in University Assessment and Feedback*, eds C. Kreber, C. Anderson, N. Entwistle, and J. McArthur (Edinburgh: University of Edinburgh), 13–31.
- Brancaleone, D., and O'Brien, S. (2011). Educational commodification and the (economic) sign value of learning outcomes. *Br. J. Sociol. Educ.* 32, 501–519. doi: 10.1080/01425692.2011.578435
- Carless, D., Chan, K. K. H., To, J., Lo, M., and Barrett, E. (2018). "Developing students' capacities for evaluative judgment through analysing exemplars," in *Developing Evaluative Judgment in Higher Education: Assessment for Knowing and Producing Quality Work*, eds D. Boud, R. Ajjawi, P. Dawson and J. Tai (London: Routledge), 108–116.
- Dawson, P. (2015). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assess. Eval. High. Educ.* 42, 347–360. doi: 10.1080/02602938.2015.1111294
- Frail, J., Panadero, E., and Pardo, R. (2017). Co-creating rubrics: the effects on self-regulated learning, self-efficacy and performance of establishing assessment criteria with students. *Stud. Educ. Eval.* 53, 69–76. doi: 10.1016/j.stueduc.2017.03.003
- Hack, C. (2015). Analytical rubrics in higher education: a repository of empirical data. *Br. J. Educ. Technol.* 46, 924–927. doi: 10.1111/bjet.12304
- Hack, C., Devlin, J., and Lowery, C. (2014). Analytical rubrics: a repository of empirical data. doi: 10.13140/2.1.4647.7766
- Hofer, B. K. (2001). Personal epistemology research: implications for learning and teaching. *Educ. Psychol. Rev.* 13, 353–383. doi: 10.1023/A:1011965830686
- Hudson, J., Bloxham, S., den Outer, B., and Price, M. (2017). Conceptual acrobatics: talking about assessment standards in the transparency era. *Stud. High. Educ.* 42, 1309–1323. doi: 10.1080/03075079.2015.1092130
- Jackel, B., Pearce, J., Radloff, A., and Edwards, D. (2017). Assessment and feedback in higher education: a review of literature for the higher education academy.
- Jankowski, N., and Proveis, S. (2014). Neoliberal ideologies, governmentality and the academy: an examination of accountability through assessment and transparency. *Educ. Philos. Theory* 46, 475–487. doi: 10.1080/00131857.2012.721736
- Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assess. Eval. High. Educ.* 39, 840–852. doi: 10.1080/02602938.2013.875117
- Kruger, J., and Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J. Pers. Soc. Psychol.* 77, 1121–1134. doi: 10.1037/0022-3514.77.6.1121
- Norton, L. (2004). Using assessment criteria as learning criteria: a case study in psychology. *Assess. Eval. High. Educ.* 29, 687–702. doi: 10.1080/0260293042000227236
- O'Donovan, B., Price, M., and Rust, C. (2004). Know what I mean? Enhancing student understanding of assessment standards and criteria. *Teach. High. Educ.* 9, 325–335. doi: 10.1080/1356251042000216642
- Orr, S. (2007). Assessment moderation: constructing the marks and constructing the students. *Assess. Eval. High. Educ.* 32, 645–656. doi: 10.1080/02602930601117068
- Panadero, E., and Jonsson, 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
- Raaper, R. (2016). Academic perceptions of higher education assessment processes in neoliberal academia. *Crit. Stud. Educ.* 57, 175–190. doi: 10.1080/17508487.2015.1019901
- Reddy, Y. M., and Andrade, H. (2010). A review of rubric use in higher education. *Assess. Eval. High. Educ.* 35, 435–448. doi: 10.1080/02602930902862859
- Rodríguez-Gómez, G., and Ibarra-Sáiz, M. S. (2015). "Assessment as learning and empowerment: towards sustainable learning in higher education," in *Sustainable Learning in Higher Education*, eds M. Peris-Ortiz and J. M. Merigo Lindahl (Springer), 1–20.
- Sadler, D. R. (2007). Perils in the meticulous specification of goals and assessment criteria. *Assess. Educ. Princip. Policy Pract.* 14, 387–392. doi: 10.1080/09695940701592097
- Sadler, D. R. (2009). Grade integrity and the representation of academic achievement. *Stud. High. Educ.* 34, 807–826. doi: 10.1080/03075070802706553
- Strathern, M. (2000). The tyranny of transparency. *Br. Educ. Res. J.* 26, 309–321. doi: 10.1080/713651562
- 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. Princip. Policy Pract.* 14, 281–294. doi: 10.1080/09695940701591867
- Tummons, J. (2014). The textual representation of professionalism: problematising professional standards for teachers in the UK lifelong learning sector. *Res. Post Compul. Educ.* 19, 33–44. doi: 10.1080/13596748.2014.872918

**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 Bearman and Ajjawi. 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.





# 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 indeterminateness 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.

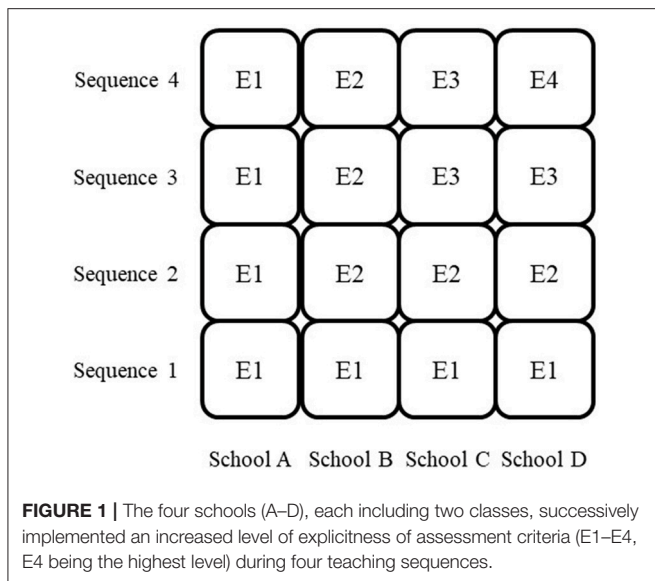
These findings are indeed indications of positive effects on students' SRL, but the students using rubrics in the study by Panadero and Romero also reported higher levels of stress while performing the task as compared to the control group. Furthermore, the learning-oriented SRL scores decreased for 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.

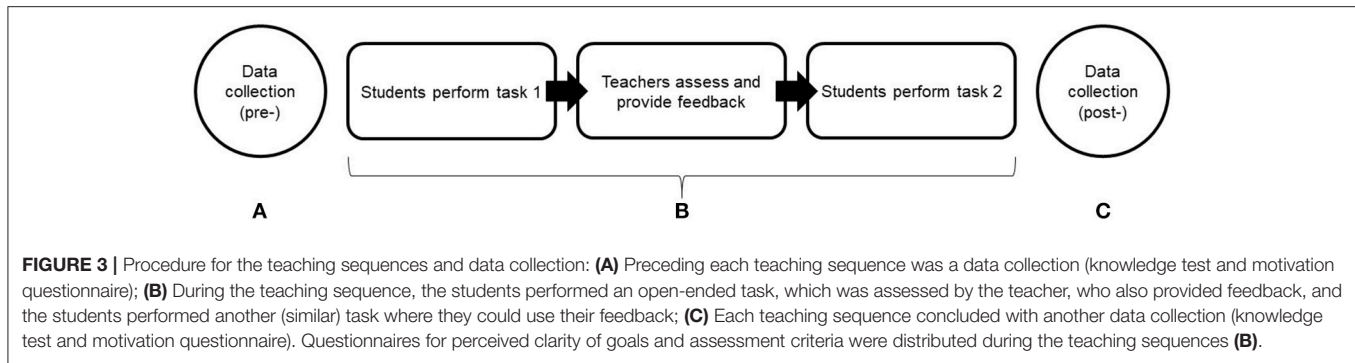
delivered orally to students, either individually, in pairs, or in small groups, depending on how the teachers arranged this purely formative event. Students then performed a similar task (or revised the first one), as an incentive to actively make use of the criteria (**Figure 3**). It is important 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<br>(13.3) | 64.3<br>(19.0) | 79.8<br>(16.6)  | 70.7<br>(16.4) |
| Post-test | 81.4<br>(29.5) | 87.6<br>(17.9) | 105.8<br>(21.0) | 87.1<br>(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 Srijbos, 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.



# Learning to See New Things: Using Criteria to Support Pre-service Teachers' Discernment in the Context of Teachers' Relational Work

Pernilla Holmstedt\*, Anders Jönsson and Jonas Aspelin

Faculty of Education, Kristianstad University, Kristianstad, Sweden

## OPEN ACCESS

### Edited by:

Mary Frances Hill,  
University of Auckland, New Zealand

### Reviewed by:

David Alexander Berg,  
University of Otago, New Zealand  
Quincy Luciani Elvira,  
Radboud University Nijmegen,  
Netherlands

Susan M. Brookhart,  
Duquesne University, United States

### \*Correspondence:

Pernilla Holmstedt  
pernilla.holmstedt@hkr.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:** 19 June 2018

**Published:** 09 July 2018

### Citation:

Holmstedt P, Jönsson A and Aspelin J  
(2018) Learning to See New Things:  
Using Criteria to Support Pre-service  
Teachers' Discernment in the Context  
of Teachers' Relational Work.  
Front. Educ. 3:54.  
doi: 10.3389/feduc.2018.00054

The purpose of this study was to investigate how pre-service teachers' understanding of relational competence can be supported through the use of digital video and explicit criteria. The study is a mixed method intervention study, where pre-service teachers analyzed the teacher-student relationship as depicted in a short video sequence with the support of explicit criteria. These analyses were analyzed with content analysis according to the criteria and a thematic comparison of pre-service teachers' analyses before and after the access to explicit criteria. Findings suggest that the use of explicit criteria supported pre-service teachers' discernment of significant dimensions of teacher-student relationships, so that they were able to discern and discuss aspects of the teacher-student relationship with a specific focus on teacher-student interaction and with greater detail and nuance. The study also provides some tentative evidence that modeling the use of criteria may support pre-service teachers' use of the criteria.

**Keywords:** assessment, criteria, pre-service teachers, relational competency, transparency

## INTRODUCTION

During the last three decades, extensive international research, including research reviews, and meta-analyses, has shown that supportive relationships between teachers and students have beneficial effects on factors such as students' subject-specific performance, social development, satisfaction, well-being, and motivation to learn (e.g., Wubbels and Brekelmans, 2005; Cornelius-White, 2007; Hattie, 2009; Roorda et al., 2011; Sabol and Pianta, 2012; Wubbels et al., 2012). In a summary of research, Hughes (2012) claims that: "we know enough to apply the knowledge gained to the task of increasing teachers' abilities to provide positive social and emotional learning environments" (p. 319). It is not until the last decade or so, however, that researchers have implemented professional development interventions focusing on teacher-child relationships (Sabol and Pianta, 2012). Moreover, although it has been suggested that pre-service training should be a prime target for informing teachers on practices associated with high quality relationships (Sabol and Pianta, 2012), research into relational competence in teacher-education programs is largely lacking (Rimm-Kaufman et al., 2003; Nordenbo et al., 2008; Sabol and Pianta, 2012). This lack of research has made it difficult for educators to work systematically to develop teacher-student relational competence. The study reported here, aims to address this scarcity in contemporary educational research by investigating the development of pre-service teachers' understanding of

relational competence<sup>1</sup>, through the use of digital video and explicit criteria.

## BACKGROUND

The research presented here is based on the assumption that social relationships cannot be considered one factor among others; instead, all types of educational phenomena are fundamentally relational and the teacher-student relationship is the central factor underlying learning and development among students (Aspelin, 2012). In simple terms, the relational competence of teachers represents the capability to develop positive (supportive, caring, trusting, etc.) relationships with students and other significant parties. This definition implies that relational competence does not pertain to relationships in general, but rather to a certain type—educational relationships. Thus, in order for teacher interactions to meet the criteria for relational competence they must be relevant to the aims of the education. This preliminary definition serves as the point of departure for the research project presented in this article.

### Relational Competence in Teacher Education

As mentioned above, there is a general lack of research investigating relational competence in teacher-education programs. However, during the past few years, in Scandinavia, at least two such research projects have been initiated: one in Denmark and one in Sweden.

The purpose of the four-year Danish project, which involved two groups of pre-service teachers, 14 instructors, and 18 elementary school teachers, was to explore and develop the relational competence of pre-service teachers by training “attentive presence and empathy as components of relational competence” through the use of various mental and communication exercises (Skibsted and Matthiesen, 2016, p. 14, our transl.). The project was based on the idea that in order for the students to become effective teachers, they must first become well-versed in their own reactions and relationships (Skibsted and Matthiesen, 2016).

The project could be considered successful in some respects. For example, pre-service teachers who participated in the project to a greater degree than other pre-service teachers developed “a reflective and open mindful approach to their own experiences and reactions as well as to their communication with the students” (Nielsen and Fibaek Laursen, 2016, p. 43, our transl.). In addition, findings suggest that the project influenced the pre-service teachers by changing their pursuit of relational competence from a “hoping for luck” approach to one that is “reflected and intentional” (Nielsen and Fibaek Laursen, 2016, our transl.).

<sup>1</sup> The concept of relational competence is rarely used in the international discourse; terms such as “interpersonal knowledge,” “interpersonal skills,” and “social skills” are more common. The discourse on relational competence in Scandinavia is distinguished by a focus on operationalization, e.g., on how to strengthen teachers’ competence with support from different methods (Klinge, 2016). Today, relational competence is considered to be a central concept within Danish school and teacher education (Skibsted and Matthiesen, 2016).

However, there are also critical views. Matthiesen and Gottlieb (2016) hold that pre-service teachers in the project tended to use exercises and relational competence “functionally”—as tools for solving concrete problems in the classroom—but that these exercises did not lead to any deeper pedagogical insights. In addition, they argue that training in this subject is mainly directed toward the “reflective domain,” with a focus on the pre-service teachers’ understanding of themselves rather than on their relationships with the students and that the actual relational competence of these pre-service teachers has been neglected. Matthiesen (2016) expounds on this criticism and holds that the concept of relational competence, as used in this project, was tantamount to an individualized rationality that urged pre-service teachers “to gaze inward rather than outward in the relationship” (our transl.). As an alternative, Matthiesen champions a “relational judgment discourse” in which the teacher responds judiciously when engaged in particular social interactions and, not least, when confronted with unfamiliar situations.

The Swedish project (performed by the authors), on the other hand, was formulated beyond the individualized rationality that Matthiesen (2016) criticizes. The attention was directed “outwards,” toward interpersonal communication between teacher and student, rather than “inwards,” toward self-reflection by teachers/pre-service teachers.

The project was a small scale pilot study, involving six pre-service teachers, using digital video to investigate how pre-service teachers responded to challenging and unpredictable situations that were “relationally problematic.” The pre-service teachers watched short video sequences and were then asked to:

1. Describe the situation: What do you notice?
2. Analyse the teacher-student relationship: (a) In what way to you think the teacher acts to *support* a positive relationship with the students?; and (b) In what way to you think the teacher *counteracts* a positive relationship with the students?
3. Describe how you think the teacher should handle the situation?

Two major themes were found in the analyses made by the pre-service teachers. First, the analyses of the teacher-student relationship were mostly general and abstract, rather than being nuanced and detailed. Second, the analyses held a view of relational competence as a type of craftsmanship and social engineering. According to such a view, the teacher is someone who designs and maintains relationships, rather than being involved in them. Common to both themes was that they referred to relatively static frameworks, general situations, and conveyance of bodies of knowledge; instead of paying attention to specific actions, or to spatial and temporal contexts and situations. To put it simply, according to the analyses, the pre-service teachers had difficulties in: (a) discerning or describing important aspects of the teacher-student relationship as displayed in the movies, and (b) analyzing the specific situations from a relational perspective. Furthermore, suggested strategies for handling the situations were primarily based on either a didactic or a leadership perspective.

## Supporting Professional Learning

Provided that pre-service teachers have difficulties in discerning important aspects of the teacher-student relationship and analyzing situations from a relational perspective, how can teacher education support pre-service teachers in developing such a relational competence? On the one hand, it has been argued that participation in a “community of practice” and non-formal learning are primary routes for learning to become a professional; starting as a peripheral participant and slowly advancing toward a more central position (Schön, 1983; Lave and Wenger, 1991; Wenger, 1998). In line with this argument, the best way to educate teachers would be to let pre-service teachers act as teachers among experienced professionals. This view is often supported by in-service teachers, as they typically claim that the main way of learning to teach is by doing the job (Metcalf et al., 1996; Knight et al., 2006).

There are, however, some potential drawbacks. For one thing, apprenticeship and non-formal learning are typically more time consuming than formal instruction. Another, and perhaps more serious point, is that workplace-based training might promote socialization into an unwanted occupational culture and outdated practices (Elliott, 1991). Furthermore, workplace-based training does not necessarily support pre-service teachers in reflecting on their practice (Metcalf et al., 1996). Acknowledging the potential drawbacks of workplace-based training, however, is not tantamount to arguing that teacher education should be entirely theoretical or entirely campus-based. Instead, it suggests a distinction as to which competencies are best learned in workplace settings and those more properly learned in other settings. This is because there seem to be limitations as to what can be learned through participation, or through more “vicarious means.” Regarding the latter, Elliott (1991) notes that even though professional learning (just like any other learning) is situated and experiential, it does not have to involve direct participation. Practical situations can also be experienced vicariously, for example by reflecting on case studies and/or discussing different ways to act in relation to simulation exercises. This means that when learning other competencies than actual teaching performance, the classroom is not necessarily the optimal setting. Instead, other settings can offer alternative ways to support the development of specific skills, which is shown by research indicating that different kinds of technology (such as tape recorders in Anderson and Freiberg, 1995, computer simulations in Yeh, 2004, and video in Yerrick et al., 2005) can provide effective support for pre-service teachers when analyzing their own, or others’, instruction. Consequently, technology supported and case-based teaching has been used to support the development of a number of different complex skills, such as “reflective ability” (Metcalf et al., 1996), analyzing classroom situations (Jönsson, 2008), and communication skills (Lucander et al., 2012). Interestingly, several of these studies have made use of explicit assessment criteria, as a means for guiding students in discerning important characteristic in complex situations.

## The Use of Explicit Criteria

According to Polanyi (1967/1983), who introduced the concept of “tacit knowledge,” all human activities, even those that are

highly theoretical or scientific, have a tacit dimension. This tacit knowledge, which is grounded in unspoken traditions and experience, provides the frames for how to interpret problems, and how to go about solving them, within a given community of practice.

Assessment criteria constitute an excellent example of such tacit knowledge, since criteria are generally grounded in unspoken traditions and experience among teachers. Using explicit criteria to communicate expectations to students is therefore often criticized, since the manifest expressions of the criteria (i.e., the words) cannot convey the full complexity of the latent criteria (e.g., Sadler, 2009, 2014). Furthermore, criteria belong to a given community of practice, which means that the meaning attached to them is not easily transferred to other contexts. In other words, in order to understand the criteria, you also have to have some familiarity with the practice to which they belong.

This understanding of criteria is also reflected in findings from empirical research, where several research reviews suggest that explicit criteria (in the form of scoring rubrics) may have the potential to promote student learning by clarifying expectations, but not without a thorough implementation (Jonsson and Svingby, 2007; Reddy and Andrade, 2010; Panadero and Jönsson, 2013; Brookhart and Chen, 2015; Brookhart, 2018). In particular, there is a distinction between school settings, which typically require more comprehensive implementations, and the higher-education context, where students are often able to use criteria productively even with very limited efforts to implement them (Panadero and Jönsson, 2013; Jonsson and Panadero, 2017; Brookhart, 2018). This could be assumed to be a result of higher-education students’ familiarity with the practice to which the criteria belong.

Panadero and Jönsson (2013) also propose that it is not the explicit criteria as such, or the criteria in isolation, that clarify expectations and promote student learning, but the explicit criteria in combination with other activities, such as feedback and/or self-, and peer-assessment. However, the criteria may support the students during these activities, by guiding their attention to important aspects of their own, or others’, performance. Specifically, the transparency provided by explicit criteria has been shown to: (a) reduce student anxiety, (b) aid the feedback process, and (c) support student self-regulation; all of which may indirectly facilitate improved student performance. Furthermore, Jonsson (2014) presents findings from different case studies in professional education, where students found explicit criteria useful for self-regulation (i.e., for planning, monitoring, and evaluating their performance). Important features for supporting students’ understanding and use of the criteria were that the criteria were: (a) closely aligned with the assignments and not too general or abstract, and (b) made accessible through explanations by the teachers, timing (i.e., access during the planning phase, before performing the assignment), and easily obtainable on paper or digitally.

In addition to research suggesting that explicit criteria may facilitate higher-education students’ learning, there are a number of critics arguing against the use of explicit criteria. In most cases, the opposition is a matter of perspective. As meritoriously



explained by Ajjawi and Bearman (2018), people may hold a representational view of criteria (and/or standards), which assumes that a criterion is an accurate and stable representation of something, and that this something is separate from the knower. Criteria, in this view, are more or less easily transferred to other contexts, since each criterion has one single meaning, which does not change in relation to the context or the person who interprets them. This is in contrast to sociocultural perspectives, in which the context and its social and cultural relations are taken into account. In such a perspective, explicit criteria are only “the tip of the iceberg,” while the greater part is tacit, residing in the practices of academic, and professional communities (O’Donovan et al., 2004).

There are, however, some issues with explicit criteria that cannot be dismissed as a matter of perspective. For instance, analytic assessments, which focus on the parts, as opposed to holistic assessments, which focus on the whole, may involve a risk of fragmentation. Sadler (2009) therefore argues against the use of analytical assessment and pre-set criteria, in favor of holistic assessment with “emergent” criteria. Emergent criteria means that assessors should not set any criteria beforehand, but address criteria that surface in the moment of assessing a particular piece of work—much like the appraisal by connoisseurs of art, wine, etc. One of Sadler’s main arguments for this approach is what he refers to as the “indeterminacy of criteria”: When breaking down holistic judgments into more or less discrete components, these components—no matter how many they are and no matter how carefully they are selected—cannot sufficiently represent the full complexity of the multi-criterion qualitative judgment made by the connoisseur. To substantiate this argument, he presents a number of observations in the way assessors approach assessment and/or grading. Most of these observations are about differences between holistic and analytic judgements, such as assessors agreeing on the overall grade/score for a particular work, but not on the level of performance for individual criteria. Sadler also notes that, in his experience, teachers generally have more confidence in their own holistic judgements as compared to analytical assessments and that global judgments are often made through the lenses of the pre-set criteria. The latter means that qualities not visible through those lenses might be filtered out and not taken into account. Instead of relying on analytic assessment and pre-set criteria as a vehicle for transparency in assessment, Sadler therefore argues that students need to develop a conceptualisation of what constitutes “quality” by continuously evaluating authentic work, without being hampered by criteria specified beforehand.

The main problem with this argument is that it is not easy for novices to know what to look for in authentic work. This is all too evident in a number of studies. An illustrative example is provided by Orsmond and Merry (1996), where students were asked to assess each other’s work. Even though all criteria were explained to the students, they were unable to recognize some of these criteria in the work by their peers. As an example, a majority of students had actually drawn a “clear and justified conclusion” (which was a criterion), but did not know it. The question of using pre-set criteria or developing a conception of quality through evaluating authentic work is therefore not a

question of either one or the other. Rather, what seems to be needed is an integration of both. Students need language (i.e., criteria) to know what to look for in authentic work, but they also need to experience authentic work in order to know how the criteria may be realized. Explicit criteria can provide a scaffolding structure for students when learning to identify indicators of quality, but like other scaffolding structures it can be disregarded if not needed and gradually phased out as the students become more independent.

## AIM AND RESEARCH QUESTIONS

The purpose of this study is to investigate how pre-service teachers’ understanding of relational competence can be supported through the use of digital video and explicit criteria. The overarching question is whether explicit criteria can be used to support the discernment of important characteristic in complex situations, provided that the criteria are used as manifest expressions of the (much wider) latent criteria, and that the criteria are contextually situated. Or, more specifically in relation to this study:

How do pre-service teachers’ analyses of teacher-student interaction, as simulated through digital video, differ before and after the introduction of explicit criteria?

## METHODOLOGY

This is an intervention study, where pre-service teachers analyzed the teacher-student relationship as depicted in a short video sequence with the support of explicit criteria. These pre-service teacher analyses were then analyzed by the researchers in order to answer the research question. The research presented belongs to the mixed method research paradigm (e.g., Johnson and Onwuegbuzie, 2004), which means that both quantitative (content analysis) and qualitative (thematic analysis) methods have been used to investigate the data, in order to provide a more comprehensive (and potentially more valid) answer to the research question.

### Sample

Participants were two groups of pre-service teachers [ $n_1 = 7$  (mean age 27 years) and  $n_2 = 10$  (mean age 29)] attending a teacher-education program for teaching in grade 4–6 (i.e., students 10–12 years). The study was performed during the sixth semester of the program (the entire program is eight semesters), when the pre-service teachers attended courses on the professional work of teachers, where the focus of the study could connect to existing learning objectives. All students participated in the study, which means that the low number of participants is an effect of the low number of students attending these courses.

### Procedure

The intervention can be divided into three distinct steps:

1. The pre-service teachers watched a short video sequence, focusing on teacher-student interactions, where the teacher’s relational competency was challenged (**Figure 1**). The movie

was recorded by professional film-makers, in order to make it feel authentic and encourage the pre-service teachers to engage with the situation. The pre-service teachers analyzed the situation, using the same questions as in the pilot study described above:

- a. Describe the situation: What do you notice?
  - b. Analyse the teacher-student relationship: a) In what way to you think the teacher acts to *support* a positive relationship with the students?; and b) In what way to you think the teacher *counteracts* a positive relationship with the students?
  - c. Describe how you think the teacher should handle the situation?
2. The pre-service teachers were given access to explicit criteria for relational competency. The meaning of the criteria were explained to the pre-service teachers by an expert in relational pedagogy.
  3. The pre-service teachers analyzed the video sequence once more, with the support of the criteria.

All three steps were similar for both groups ( $n_1$  and  $n_2$ ), with one exception. For group  $n_1$  the criteria were introduced only orally, but for group  $n_2$  the expert on relational pedagogy also modeled how to use the criteria by analyzing a short sequence of the commercial movie “Precious” (directed by Lee Daniels, starring Gabourey Sidibe). This was done to acknowledge that the criteria are contextually situated and the need for students to be familiar with the practice to which the criteria belong.

## Data and Analysis

Data for this study is pre-service teachers’ written analyzes of teacher-student interactions, simulated through digital video before and after the access to explicit criteria about teachers’ relational competency. The responses by the pre-service teachers were analyzed with both quantitative content analysis and qualitative thematic analysis (Braun and Clarke, 2006; Vaismoradi et al., 2013), and the responses before and after the access to explicit criteria were compared in order to identify differences between the two occasions. The content analysis focused on the frequency of pre-service teachers’ use of concepts important to analyze the situation from a relational perspective, using the criteria as analytical tools. The frequencies were compared before and after the access to criteria, but no statistical analyses have been made, due to the small number of participants taking part in the study. Furthermore, the groups have been analyzed separately, since the intervention was slightly different in the groups, but no conclusions will be drawn based on the differences between the groups, again due to the low number of participants.

In addition to the content analysis, a thematic comparison has been made using the entire material before and after the access to criteria. This analysis is based on repeated reading of the respondents’ analyses in search of themes transcending the material. The analysis followed the procedure outlined by Braun and Clarke (2006), which, in this case, means that the following steps were taken:

1. The first step was to read through the pre-service teachers’ analyses and note initial ideas.
2. Interesting features of the data were coded across the data set.
3. Codes were assembled into initial themes, gathering data relevant to each initial theme.
4. Themes were checked against coded extracts and the data set as a whole.
5. The specifics of each theme were refined.
6. A selection of compelling extract examples for this article was made.
7. The extracts were translated to English by the authors.

## Criteria for Relational Competency

In this study, Scheff’s (1990) theory about social bonds was used in order to formulate explicit criteria for teachers’ relational work<sup>2</sup>. The most central concept of the theory is the “social bond,” which, simply stated, can be defined as the forces that hold people and groups in the community together. Although these bonds between people may appear well established and lasting, in reality they are temporary, dynamic, and unpredictable. You can therefore never be completely sure that relationships will have a certain character and social bonds are more or less constantly tested. The quality of social bonds ranges from fragile and uncertain to strong and secure. The bonds can be built, repaired, threatened, or even cut-off. What is crucial for the quality of the bonds is how participants communicate with each other and how well they are “attuned.” “Attunement” refers to people’s cognitive and emotional adjustment to each other in the interpersonal communication, both verbal (what is being said) and non-verbal (how it is said and expressed). The degree of attunement depends on how well individuals understand each other and the extent to which they show each other adequate and due respect.

Another concept is “differentiation,” which refers to the degree of closeness and distance in interpersonal relations. Scheff assumes that differentiation is a fundamental dilemma in human relationships. When two people become so close that they can experience each other’s side of the relationship, yet are distanced enough from each other that they perceive themselves as unique, individual entities, we can speak of optimal differentiation. Neither individual components nor social components are overemphasized in such a relationship; instead a balance is achieved between closeness and distance. However, should one or the other, or both parties, experience excessive distance—that is, if direct contact with the other is absent and the importance of the self is overemphasized—we can speak of *over-differentiation* or *isolation*. Similarly, when individuals experience excessive closeness—lose contact with vital aspects of themselves and when the importance of the other person/group is overemphasized—we can speak of *under-differentiation* or *engulfment*.

Emotions also play a vital role in Scheff’s theory. Stable social bonds imply lasting and relatively deep emotional connections and Scheff defines shame and pride as fundamental social

<sup>2</sup>Scheff (1990) is an American social psychologist/sociologist and his book *Microsociology* is by many considered to be his magnum opus (on which his later works are based). With some notable exceptions (Aspelin, 2006, 2010; Beaulieu, 2016), his theory has rarely been applied to the educational context and, more specifically, to the teacher-student relationship.

Parent: Hi.

Teacher: Hi. Hello Johanna.

Student: Hi.

Teacher: Then let's see. I have received assessments from your other teachers, in other subjects. Then let's see ... eeh. First we have art, that's Marianne, and she writes that you are very good in arts, thorough and take the assignments seriously. But she also writes that you do not say so much in the classroom when you have class discussions and things like that, and that is something you could improve. Do you recognize this?

Student: Uhm.

Teacher: Uhm, then we have physical education, that's Birgitta. Physical education is also looking good, it says that you have a D, but a C or a B at several assignments. That seems good, uhm? It is says that you could have a higher grade if you made spoke up sometimes, if you took a little more initiative, if you were a bit more engaged, uhm. And then we have craft and design, I can see that you've had woodwork this semester. That's Annika. Also good reviews for what you have performed, but under improvements it says that you are very quiet during class discussions and ...

Parent: Now you've got to stop goddammit!

Teacher: What?

Parent: You sit here talking about the same god damn thing over and over again ... that she's quiet. What has that got to do with her school work?

Teacher: Yes, but we have to ...

Parent: Then why are you sitting here nagging about the same thing all over again, that she is quiet? That's who she is, when will you accept that? She is good at what she is doing, is she not? Or what the hell are you looking for really, that everyone should be the same?

**FIGURE 1** | Transcript from the movie that the preservice teachers analyzed.

emotions. Shame and pride are awakened in a context where the individual visualizes how he/she behaves and is valued in the eyes of the other. Positive role-taking is initiated by and leads to feelings of pride, while negative role-taking is associated with feelings of shame. Therefore stable bonds are signaled by feelings of pride and unstable bonds by feelings of shame. Shame and pride are technical terms and umbrella concepts for a range of

emotions within each group. These emotions are not viewed as being inherently positive or negative, but rather as messengers reflecting the qualities of interpersonal relationships.

With the aid of Scheff's theory, a more nuanced description can be made of teachers' relational competency. Scheff holds that attunement is crucial for understanding the quality of the social bond in interpersonal communication. Relationally competent

teachers therefore need to communicate in such a way that they and the students form strong social bonds with each other. As we have seen, this requires mutual understanding and respect. Consequently, teachers need to make themselves understood and understand—and demonstrate that they understand—the students. Teachers also need to show respect for students while acting in a way that promotes students' respect for them. This first aspect of relational competence will be called *communicative competence* and reflects the ability of teachers to communicate both verbally and non-verbally in order to achieve a high degree of cognitive and emotional attunement in relation to students. In this regard, the actions of a relationally competent teacher encourage mutual understanding and respect in the work with students.

The second aspect of relational competency is *differentiation competence*, which reflects the ability of teachers to act in such a way that neither they nor the students become too close nor too distant from each other. A relationally competent teacher acts in a way that space is created to allow both students and teachers to discern themselves as individuals, without jeopardizing social bonds.

*Socio-emotional competence* is the third aspect of relational competency and this concept reflects the importance of teachers' attunement toward emotional signals in interpersonal communication. A relationally competent teacher acts in order to evoke and encourage feelings of pride, while acknowledging and channeling feelings of shame in a direction that is productive from the standpoint of educational goals.

From the three aspects of relational competency described above, criteria relating to communication, differentiation, and emotions were formulated and shared with the pre-service teachers. Furthermore, an additional criterion, focusing on teachers' professional work was added to the framework, since this was the specific content of the courses they attended. The criterion "Professionalism" reflects whether the teacher acts in a way that can be expected from a professional who is accountable for her actions. All criteria can be found in Appendix.

## FINDINGS

In this section, the content analyses (before and after the access to criteria) are presented first, then the thematic comparison. In order not to confuse the pre-service teachers with the teacher or student in the analyses, the pre-service teachers are called "respondents" in this section. The individual respondents are identifiable by letters A-Q and all quotes have been translated from Swedish by the authors.

### Content Analysis Before the Access to Explicit Criteria

In relation to the *communication* criterion, all respondents discussed the verbal communication. The majority of the respondents focused on how the conversation was organized, for instance that the teacher was the one speaking and that the teacher did not ask any questions or invited the student into the discussion. The respondents thought that the student should

have participated in the discussion. Only one respondent made a connection between verbal communication and the purpose of understanding or being understood:

The teacher uses a language that is understandable to herself, maybe not for the student and the parent. (Respondent A, group n<sub>1</sub>)

Besides this example, there were no connections between the verbal or non-verbal communication for the purpose of understanding or being understood. Some respondents claimed that the teacher, through her verbal communication, invited the student to be part of the discussion. However, this was interpreted as a way for the teacher to make the student involved, and not for the purpose of understanding and being understood:

The teacher shows that she wants to invite the student into a dialogue when she asks whether the student agrees. (Respondent C, group n<sub>1</sub>)

.../ the teacher invites the student to the conversation by asking questions. (Respondent O, group n<sub>2</sub>)

One respondent thought that the teacher, through her non-verbal communication, invited the student to be part of the discussion. This was interpreted as a way for the teacher to make contact and not for the purpose of understanding and being understood:

.../ she [the teacher] sometimes looks up and smiles toward the student, indicating that the teacher still wants to make contact. (Respondent K, group n<sub>2</sub>)

In total, about half of the respondents discussed the non-verbal communication of the teacher, such as the way of speaking, facial expressions, and eye contact, as something that matters to the teacher-student relationship:

The teacher has a positive voice when she reviews the assessments. (Respondent D, group n<sub>1</sub>)

A smile usually smooths such nervous and tense situations. (Respondent I, group n<sub>2</sub>)

Some respondents discussed the student's non-verbal communication and thought that the student, through her non-verbal communication, clearly showed that she was uncomfortable in the situation, but that the teacher was not aware of this:

The student shows through her body language that she is uncomfortable in the situation. (Respondent M, group n<sub>2</sub>)

In relation to *differentiation*, only two respondents made reference to this criterion:

The student is not involved in the conversation and the teacher has a somewhat distant relationship with both the student and the parent. (Respondent J, group n<sub>2</sub>)



In relation to *emotions*, the majority of respondents discussed the student's feelings and thought that the situation was difficult for the student. Almost half of the respondents thought that the teacher failed to acknowledge the student's feelings. This is linked to non-verbal communication and that the student clearly displayed her feelings with her body language. None of the respondents discussed the teacher's feelings.

In relation to *professionalism*, only one respondent mentioned this aspect in the analysis:

The teacher has a professional behavior and acts as a teacher. She probably does the same with all the students and does not treat anyone differently. (Respondent E, group n<sub>1</sub>)

The majority of the respondents, however, thought that the teacher needed to respond to the students differently, which can be linked to professionalism and that the teacher acts as can be expected. Above all, didactic perspectives on what the teacher should do next predominated respondents' analyses, such as making clarifications to the student or being more dialogic in the conversation.

There were no clear differences between the groups' first analyzes in relation to the criteria Communication and Differentiation (Table 1). However, there were differences in relation to Emotions and Professionalism. Of the respondents in group n<sub>1</sub>, only about half of the respondents discussed the significance of emotions for the teacher-student relationship, as compared to 9 out of 10 in group n<sub>2</sub>. The groups also differed in terms of the extent to which they discussed the teacher's responsiveness to the students' feelings, where only 2 out of 7 discussed this in group n<sub>1</sub>, as compared to 9 out of 10 in group n<sub>2</sub>. A difference between the groups in relation to professionalism was the extent to which the respondents mentioned the teacher's response to the student, where 3 out of 6 respondents in group n<sub>1</sub> mentioned this, as compared to 9 out of 10 respondents in group n<sub>2</sub>.

**TABLE 1** | Comparison of analyses before access to explicit criteria for groups n<sub>1</sub> and n<sub>2</sub>.

|                        | Group n <sub>1</sub> (n = 7) | Group n <sub>2</sub> (n = 10) | In total (n = 17) |
|------------------------|------------------------------|-------------------------------|-------------------|
| <b>COMMUNICATION</b>   |                              |                               |                   |
| C:1                    | 1                            | –                             | 1                 |
| C:2                    | 1                            | –                             | 1                 |
| C:3                    | 4                            | 4                             | 8                 |
| <b>DIFFERENTIATION</b> |                              |                               |                   |
| D:1                    | 1                            | 1                             | 2                 |
| <b>EMOTIONS</b>        |                              |                               |                   |
| E:1                    | 3                            | 9                             | 12                |
| E:2                    | 2                            | 9                             | 11                |
| <b>PROFESSIONALISM</b> |                              |                               |                   |
| P:1                    | 1                            | –                             | 1                 |
| P:2                    | 3                            | 9                             | 12                |

## Content Analysis After the Access to Explicit Criteria

After the access to explicit criteria, all respondents discussed the verbal *communication*; that the teacher is the one speaking and that the student should have been more involved. A significant difference in this analysis, as compared to the former, was that the majority of the respondents also discussed that the communication should aim at the teacher and the student understanding each other. Some respondents also mentioned that the communication was not attuned. The majority of respondents discussed communication based on the concept of understanding, both from the perspective of the teacher and from the perspective of the student:

The teacher focuses on explaining, but not on getting the student to understand or to understand the student herself. (Respondent O, group n<sub>2</sub>)

The teacher focuses to some extent on being understood, but does not read the student's signals. The teacher could have sought to understand the pupil better. /.../ The student cannot make herself understood, since she is not given any room to speak. (Respondent K, group n<sub>2</sub>)

In addition, one respondent wrote that the teacher tried to be responsive to and build on the student's thoughts. Another one wrote that the teacher tried to follow the student's thoughts, which also connects to the purpose of understanding.

The majority of respondents wrote that the teacher confirmed, or did not confirm, the student through her verbal communication:

The teacher confirms the student's presence by saying her name. (Respondent M, group n<sub>2</sub>)

The teacher should have invited the student more, to confirm that she is important. (Respondent C, group n<sub>1</sub>)

Almost half of the respondents also discussed that the teacher and/or the parent/student confirmed, or did not confirm, each other through non-verbal communication:

The teacher confirms the student by looking up sometimes when she talks with the student. (Respondent O, group n<sub>2</sub>)

The teacher and the parent look at each other with small nodding confirmations. (Respondent J, group n<sub>2</sub>)

The teacher does not confirm the student's obvious body language, showing that the student is not comfortable in the situation. (Respondent Q, group n<sub>2</sub>)

In the first analysis, only a few respondents discussed the teacher's non-verbal communication as gestures, ways of speaking, facial expressions, body position, eye contact, etc., as something that mattered to the teachers-student relationship. In the second analysis, the majority of respondents discussed the non-verbal

communication of the teacher as something that is relevant to the teacher-student relationship:

The teacher's body language is unsympathetic and distant rather than inviting. (Respondent L, group n<sub>2</sub>)

Also, she is not dialogic in her body language, she does not invite either the mother or the student with her body language. (Respondent L, group n<sub>2</sub>)

Unlike the first analysis, where only one respondent discussed *differentiation* in his/her analysis, the majority of respondents discussed this as something that is relevant to the teacher-student relationship in the second analysis:

The teacher switches between closeness and distance in a way that is not particularly appropriate. The student is also distant by turning her eyes and collapsing into the chair. The parent also moves closer to her daughter, which gives a protective feeling at the same time as she moves away from the teacher. (Respondent Q, group n<sub>2</sub>)

/.../ neither the teacher nor the student tries to approach each other. (Respondent P, group n<sub>2</sub>)

The teacher shows distance through physical placement at the teacher's desk. (Respondent M, group n<sub>2</sub>)

The majority of respondents discussed the student's *emotions*. The respondents thought that the teacher was not sensitive to the student's feelings. In the first analysis, none of the respondents discussed the teacher's emotions, but in the second analysis, almost half of the respondents discussed how the teacher managed her own feelings:

The situation could have been different if the teacher was able to control her feelings. (Respondent F, group n<sub>2</sub>)

The teacher cannot handle her own feelings. (Respondent O, group n<sub>2</sub>)

Unlike the first analysis, when only one respondent mentioned *professionalism* in his/her analysis, all respondents discussed the teacher's actions from this perspective in the second analysis. However, the respondents focused on different aspects. Didactic perspectives were still discussed, but not to the same extent as in the first analysis. Other perspectives on professionalism dominated. Several respondents connected professionalism to accountability:

The teacher acts irresponsibly, probably unwittingly. (Respondent D, group n<sub>1</sub>)

The teacher tries to avoid taking responsibility for the student. (Respondent J, group n<sub>2</sub>)

A couple of respondents made connections to the professional ethics of teachers:

The teacher does not act responsibly, because she does not see the student, which is among the most important things in the profession. (Respondent K, group n<sub>2</sub>)

Several respondents suggested that the teacher was lacking in communicative competence:

Her actions are not really professional when she gets steamrolled by the mother. (Respondent I, group n<sub>2</sub>)

The parent feels forced to take over the conversation from the teacher when the misunderstandings, distances, and feelings go too far from what can be expected during a discussion between teacher, parent, and student on progress in school. (Respondent Q, group n<sub>2</sub>)

Respondents also associated professionalism with other relationship-theory concepts, such as how the teacher managed her emotions:

One should think professionally and then you should not be annoyed, but try to ignore it. (Respondent G, group n<sub>1</sub>)

She cannot handle the parent's annoyance appropriately, but immediately begins to defend herself. (Respondent O, group n<sub>2</sub>)

Overall, only one respondent did not link professionalism to the teacher's response toward the student.

In the second analysis there were major differences between the groups in terms of how they discussed relational competence based on the concepts of communication and differentiation. Nine out of 10 respondents in group n<sub>2</sub> gave examples of how the communication aimed at the teacher and the student understanding each other. In group n<sub>1</sub>, only half of the respondents (4/7) discussed this. All respondents in group n<sub>2</sub> discussed the importance of the teacher's non-verbal communication, such as gestures, way of speaking, facial expressions, body positioning, eye contact etc., as compared to five out of seven in group n<sub>1</sub>. Furthermore, all respondents in group n<sub>2</sub> provided examples of how issues of closeness and distance can be important for the teacher-student relationship, as compared to half of the respondents in group n<sub>1</sub>. The difference remained between the groups in terms of how they discussed the importance of emotions in the teacher-student relationship. All respondents in group n<sub>2</sub> gave examples of how emotions may be of significance for the teacher-student relationship, while about half of the respondents in group n<sub>1</sub> discussed this. Only one respondent in group n<sub>1</sub> discussed the teacher's responsiveness to the student's feelings, while all respondents in group n<sub>2</sub> discussed this. However, this difference was present already in the first analysis and no clear differences between how the groups discussed professionalism can be distinguished. The use of relational concepts in both groups, and both before and after the access to criteria, is summarized in **Table 2**.

**TABLE 2 |** Comparison of analyses before and after access to explicit criteria for groups  $n_1$  and  $n_2$ .

|                        | Group $n_1$ ( $n = 7$ ) |       | Group $n_2$ ( $n = 10$ ) |       | In total ( $n = 17$ ) |       |
|------------------------|-------------------------|-------|--------------------------|-------|-----------------------|-------|
|                        | Before                  | After | Before                   | After | Before                | After |
| <b>COMMUNICATION</b>   |                         |       |                          |       |                       |       |
| C:1                    | 1                       | 4     | –                        | 9     | 1                     | 13    |
| C:2                    | 1                       | 3     | –                        | 8     | 1                     | 11    |
| C:3                    | 4                       | 5     | 4                        | 10    | 8                     | 15    |
| <b>DIFFERENTIATION</b> |                         |       |                          |       |                       |       |
| D:1                    | 1                       | 4     | 1                        | 10    | 2                     | 14    |
| <b>EMOTIONS</b>        |                         |       |                          |       |                       |       |
| E:1                    | 3                       | 4     | 9                        | 10    | 12                    | 14    |
| E:2                    | 2                       | 1     | 9                        | 10    | 11                    | 11    |
| <b>PROFESSIONALISM</b> |                         |       |                          |       |                       |       |
| P:1                    | 1                       | 6     | –                        | 10    | 1                     | 16    |
| P:2                    | 3                       | 6     | 9                        | 9     | 12                    | 15    |

## Thematic Comparison of Analyses Before and After Access to Criteria

### A Change in Focus

In the first analysis, the respondents had didactic aspects in focus when analyzing the situation, despite the fact that they had been explicitly instructed to focus on teacher-student relationship. In particular, the respondents focused on how the teacher communicated the assessments. For example, one respondent wrote:

Regarding the assessment, as mentioned by the teacher, I think it is remarkable that the student has a D in physical education, but has performed at levels C and B at several occasions. It should not be possible to perform at a level B on individual assignments?<sup>3</sup> And I don't think that summative assessments belong in a conversation on student's progress, it should focus on the student's opportunities for further development. (Respondent Q, group  $n_2$ ).

This respondent emphasized that the teacher was not sufficiently prepared for the meeting, that she did not perform the conversation appropriately, and that there were shortcomings in the teacher's assessment practice. With some variations, this pattern was repeated in all analyses, as illustrated by the following citations:

The subjects that the teacher focuses on in the movie are definitely not the ones where discussion is needed, as in civics for example, but not even there I think there is a need to be able to discuss things orally. (Respondent L, group  $n_2$ ).

<sup>3</sup>In the Swedish grading system, all grades are composite measures and as such do not apply for individual assignments. However, while there are explicit requirements for levels A, C, and E in the national curriculum, grades B and D lack such requirements and are used only as intermediate grades between A-C and D-E respectively.

The students should not be assessed by how silent she is, but on how she performs in the classroom and, as said, here the focus is on the silence /.../ The situation can be amended by the teacher explaining better what the teachers mean when they "complain" about the student's silence /.../ (Respondent I, group  $n_2$ ).

The focus of the respondents changed significantly in the second analysis. In the analyses they wrote, the interaction between teacher and student was perceived as the central theme of the situation. For example, respondent Q wrote:

The teacher turns her eyes to the student, but still focuses on her papers as she talks. The teacher also hides her body language behind the table and the pen /.../ The distance from the teacher to the student is closer in the beginning, but is increasing more and more during the conversation. /.../ It is clear that the student is not comfortable with the situation, but despite this, the teacher continues the conversation as if the student does not show anything. /.../ The student shows "hiding behavior", as she looks down at the table trying to hide her face behind her hands on several occasions. (Respondent Q, group  $n_2$ )

Respondent Q focused on describing and interpreting how the teacher and the student behaved, as well as what it meant for their relationship. This pattern permeated respondents' analyses, which is illustrated by the following citations:

However, she [the teacher] does not confirm the student /.../ because she does not see the student giving signals that she is anxious about being silent. The teacher cannot see this since she concentrates on presenting the assessments while the student is quietly staring at the table. (Respondent K, group  $n_2$ )

The student looks down at the bench, holds her head and hides her eyes. The teacher sees that the student is not comfortable but does not pay attention to it. It looks like she tries to escape by continuing to talk about the assessments. (Respondent B, group  $n_2$ )

## A More Specific and Nuanced Way of Understanding Relationships

Most respondents wrote about the teacher-student relationship already during the first analysis. However, the respondents' formulations were comparatively simple and general:

However, the students seem to think that the situation and the assessments were uncomfortable. (Respondent Q, group  $n_2$ )

The teacher needs to change her entire attitude and, above all, create a better relationship with the student. (Respondent L, group  $n_2$ ).

The teacher is quite straightforward and does not seem to take into consideration that conversations about student progress can be uncomfortable for the student and she does nothing to make the situation easier. (Respondent I, group  $n_2$ )

The respondents used quite unspecific expressions, such as "the situation is uncomfortable" and that the teacher "needs to create

a better relationship.” In comparison, their descriptions and interpretations of relationships during the second analysis were comparatively specific and nuanced:

There is no closeness what so ever, although the teacher tries to create some when she asks questions like “recognize this?” and “that seems good?” but there are no questions that the student seems to want to answer or is given the opportunity to answer, because of the way the questions are asked. (Respondent I, group  $n_2$ ).

I think the teacher is too distant from the student. Partly, she does not ask how the student experiences the conversation, and partly she does not read the student’s body language, which means that she does not notice that the student is feeling uncomfortable about the conversation. (Respondent K, group  $n_2$ ).

The communication during the conversation is one-sided. The teacher is the one speaking while the student answers with “uhm” /.../ The way the teacher talks is disrespectful, according to me. She does not look at the student very much. (Respondent B, group  $n_2$ )

The thematic comparison shows that respondents’ descriptions and analyzes significantly changed during the intervention. The change consisted, first of all, of a shift in focus. In the first analysis the respondents focused on questions relating to the organization and execution of the conversation and the teacher’s assessment practice. In the second analysis, they focused on the interaction between the teacher and the student. The second change was from a comparatively general and simplistic way to analyze the situation to a more specific and nuanced way.

## DISCUSSION

This study aimed to investigate how pre-service teachers’ understanding of relational competence can be supported through the use of digital video and explicit criteria. In order to investigate this, pre-service teachers’ analyses of a simulated situation were analyzed with content and thematic analyses, both before and after the access to explicit criteria. As indicated by the findings presented above, there has been a quantitative as well as a qualitative change in the analyses made by the respondents. The content analysis clearly shows that references were made much more frequently by the respondents to important dimensions of the teacher-student relationship when having access to explicit criteria. This is true for all criteria, except for *Emotions*, to which respondents in group  $n_2$  made frequent references already before access to the criteria. Furthermore, the thematic comparison suggests that respondents’ analyses are characterized by a change of focus (from organization to interaction), as well as becoming more detailed and specific. These findings can be interpreted as the respondents’ discernment of significant dimensions of the teacher-student relationships has been affected by the use of explicit criteria, so that they—with the aid of the criteria—may see and analyze aspects of the situation that they did not notice without them.

The findings from this study thus corroborate previous research on the use of criteria, reporting that higher-education students are often able to use criteria productively even with very limited efforts to implement them (Panadero and Jönsson, 2013; Jönsson and Panadero, 2017; Brookhart, 2018). Although the pre-service teachers in this study were not familiar with the specifics of relational competency, they were familiar with other areas of teachers’ professional work, which could have facilitated the interpretation of the criteria. It should be noted, however, that the greatest changes occurred in group  $n_2$ , where the use of the criteria was modeled by an expert. Unfortunately, due to the small number of participants, it is not possible to compare the groups statistically, which means that it cannot be excluded that the observed difference may be a result of chance alone. Still, it is a reasonable assumption that the modeling supported the pre-service teachers in interpreting the criteria. Panadero and Jönsson (2013) have also proposed that it is not the explicit criteria in isolation from other activities, that clarifies expectations and promote student learning, but the combination with for instance feedback and/or self-, and peer-assessment. In this study, it was modeling that contributed to aligning the criteria with the task at hand and making them accessible (cf. Jönsson, 2014).

An alternative interpretation of the findings is that the respondents have learned to use the criteria in a mechanical/instrumental way, without a deeper understanding of relational competence. Based on the content analysis alone, no distinction could have been made between such a surface approach and a deeper understanding. However, the thematic comparison suggests a deeper understanding of the concepts used, at least in group  $n_2$ . For example, the respondents expressed themselves in a more nuanced way about both the verbal and non-verbal *communication*. They also discussed implications of *differentiation* in the teacher-student relationship and they expressed themselves in much more detail about socio-emotional aspects of the situation. The most likely explanation is therefore that the pre-service teachers have gained an understanding of how to use the concepts of relational pedagogy to analyze the situation.

Taken together, the findings from this study suggest that the use of explicit criteria supported pre-service teachers’ discernment of significant dimensions of teacher-student relationships in a simulated situation, so that they were able to discern and discuss aspects of the teacher-student relationship with another focus and with greater detail and nuance. The study also provides some tentative evidence that modeling may support pre-service teachers’ use of the criteria.

## LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

There are several important limitations of this study, which need to be kept in mind when interpreting the findings.

First, this is a small scale study with a very limited number of participants. The findings may therefore depend on the specific individuals and the findings may not necessarily generalize to



any other population of pre-service teachers, not even at the same university. Further research is thus needed in order to corroborate the findings.

Second, the focus of this study was to investigate how pre-service teachers analyze simulated situations. Consequently, no claims can be made regarding how the respondents act (or would act) in “real situations.”

Third, respondents only analyzed one simulated situation, which also limits the possibility to make any general claims about students’ proficiency in applying their knowledge about relational competence in other situations.

From the findings and limitations of this study, it is suggested that future research involves other, and larger, samples of pre-service teachers in order to substantiate the findings reported here, but also a wider spectrum of situations. It is further suggested that future research investigates to what extent pre-service teachers may apply their knowledge about relational competence in authentic settings, such as during their practicum.

## IMPLICATIONS

There are two main implications from this study. First, in line with previous research on the use of explicit criteria (e.g., Brookhart, 2018), students in higher education may use criteria productively even with relatively limited efforts of implementation. This would suggest that explicit criteria can be used in different areas, where students are in need of discerning and analyzing/evaluating complex situations.

Second, since research into relational competence in teacher-education programs is largely lacking, it is difficult for educators to design interventions to aid pre-service teachers’ development of relational competence. This study therefore makes a contribution by presenting an intervention, which has been successful in supporting pre-service teachers’ discernment

of significant dimensions of teacher-student relationships. The intervention could be used as a starting point for educators when designing other interventions, aiming to aid pre-service teachers’ development of relational competence.

## ETHICAL 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 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 consent have been given by all subjects in accordance with the Declaration of Helsinki.

## AUTHOR CONTRIBUTIONS

All authors have contributed to the design of the study, the literature review, data collection, and writing of the manuscript. The content analysis was performed by PH and the thematic comparison by JA.

## FUNDING

The research presented has been funded by Kristianstad University, Sweden.

## REFERENCES

- Ajjawi, R., and Bearman, M. (2018). “Problematising “standards”: representation or performance?,” in *Developing Evaluative Judgement in Higher education: Assessment for Knowing and Producing Quality Work*, eds D. Boud, R. Ajjawi, P. Dawson, and J. Tai (Oxon; New York, NY: Routledge), 41–50.
- Anderson, J. B., and Freiberg, H. J. (1995). Using self-assessment as a reflective tool to enhance the student teaching experience. *Teacher Educ. Q.* 22, 77–91.
- Aspelin, J. (2006). Beneath the surface of classroom interaction: reflections on the microworld of education. *Soc. Psychol. Educ.* 9, 227–244. doi: 10.1007/s11218-006-0008-2
- Aspelin, J. (2010). *Sociala relationer och pedagogiskt ansvar [Social Relationships and Educational Responsibility]*. Malmö: Gleerups.
- Aspelin, J. (2012). How do relationships influence student achievement? Understanding student performance from a general, social psychological standpoint. *Int. Stud. Soc. Educ.* 22, 41–56. doi: 10.1080/09620214.2012.680327
- Beaulieu, R. (2016). A critical discourse analysis of teacher-student relationships in a third-grade literacy lesson: dynamics of microaggression. *Cogent Educ.* 3, 1–20. doi: 10.1080/2331186X.2016.1244028
- Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 77–101. doi: 10.1191/1478088706qp063oa
- Brookhart, S. M. (2018). Appropriate criteria: key to effective rubrics. *Front. Educ.* 3:22. doi: 10.3389/feduc.2018.00022
- Brookhart, S. M., and Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educ. Rev.* 67, 343–368. doi: 10.1080/00131911.2014.929565
- Cornelius-White, J. (2007). Learner centered teacher-student relationships are effective: a meta-analysis. *Rev. Educ. Res.* 77, 113–143. doi: 10.3102/003465430298563
- Elliott, J. (1991). A model of professionalism and its implications for teacher education. *Br. Educ. Res. J.* 17, 309–318. doi: 10.1080/0141192910170402
- Hattie, J. (2009). *Visible Learning. A Synthesis of over 800 Meta-Analyses Relating to Achievement*. New York, NY: Routledge.
- Hughes, J. N. (2012). Teacher-student relationships and school adjustment: progress and remaining challenges. *Attach. Hum. Dev.* 14, 319–327. doi: 10.1080/14616734.2012.672288
- Johnson, R. B., and Onwuegbuzie, A. J. (2004). Mixed methods research: a research paradigm whose time has come. *Educ. Res.* 33, 14–26. doi: 10.3102/0013189X033007014
- Jonsson, 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. (2008). *Educative Assessment for/of Teacher Competency*. Doctoral dissertation, Malmö: Malmö University.
- Jonsson, 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
- Klinge, L. (2016). *Laerarens relationskompetence. En empirisk undersøgelse af, hvordan lærerens relationskompetence viser sig i interaktioner med elever og klasser i almenundervisningen i folkeskolen [The Teacher's Relational Competence]*. Doctoral dissertation, Copenhagen: University of Copenhagen, Denmark.
- Knight, P., Tait, J., and Yorke, M. (2006). The professional learning of teachers in higher education. *Stud. High. Educ.* 31, 319–339. doi: 10.1080/03075070600680786
- Lave, J., and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Lucander, H., Knutsson, K., Salé, H., and Jonsson, A. (2012). "I'll never forget this": evaluating a pilot workshop in effective communication. *J. Dental Educ.* 76, 1311–1316.
- Matthiesen, N. (2016). "Relationskompetencebegrebet under lup: ansvar i forudsigeligheden [The concept of relational competence]," in *Relationskompetencer i læreruddannelse og skole*, eds E. Skibsted and N. Matthiesen (Fredrikshavn: Dafolo), 139–156.
- Matthiesen, N., and Gottlieb, L. (2016). "Når man lærer at være relationskompetent – fra kursus til skolepraktik [When one is taught to be relational competent]," in *Relationskompetencer i læreruddannelse og skole*, eds E. Skibsted and N. Matthiesen (Fredrikshavn: Dafolo), 89–106.
- Metcalfe, K. K., Ronen Hammer, M. A., and Kahlich, P. A. (1996). Alternatives to field-based experiences: the comparative effects of on-campus laboratories. *Teach. Teacher Educ.* 12, 271–283. doi: 10.1016/0742-051X(95)00037-K
- Nielsen, A. M., and Fibaek Laursen, P. (2016). "Relationskompetence og klasseledelse i læreruddannelse [Relational competence and classroom management in teacher education]," in *Relationskompetencer i læreruddannelse og skole*, eds E. Skibsted and N. Matthiesen (Fredrikshavn: Dafolo), 43–70.
- Nordenbo, S. E., Sogaard Larsen, M., Tiftik, N., Wendt, R. E., and Østergaard, S. (2008). *Lærerkompetanser og elevers læring i barnehage og skole [Teacher Competences and Student Learning in Preschool and School]*. København: Danmarks Pædagogiske Universitetsforlag & Dansk Clearinghouse for Uddannelsesforskning.
- O'Donovan, B., Price, M., and Rust, C. (2004). Know what I mean? enhancing student understanding of assessment standards and criteria. *Teach. High. Educ.* 9, 325–335. doi: 10.1080/1356251042000216642
- Orsmond, P., and Merry, S. (1996). The importance of marking criteria in the use of peer assessment. *Assess. Eval. High. Educ.* 21, 239–250. doi: 10.1080/0260293960210304
- 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
- Polanyi, M. (1967/1983). *The Tacit Dimension*. Gloucester: Peter Smith.
- Reddy, Y. M., and Andrade, H. (2010). A review of rubric use in higher education. *Assess. Eval. High. Educ.* 35, 435–448. doi: 10.1080/02602930902862859
- Rimm-Kaufman, S. E., Voorhees, M. D., Snell, M. E., and La Paro, K. M. (2003). Improving the sensitivity and responsivity of preservice teachers toward young children with disabilities. *Topics Early Childhood Spec. Educ.* 23, 151–163. doi: 10.1177/02711214030230030501
- Roorda, D. L., Koomen, H. M., Spilt, J. L., and Oort, F. J. (2011). The influence of affective teacher-student relationships on students' school engagement and achievement: a meta analytic approach. *Rev. Educ. Res.* 81, 493–529. doi: 10.3102/0034654311421793
- Sabol, T. J., and Pianta, R. C. (2012). Recent trends in research on teacher-child relationships. *Attach. Hum. Dev.* 14, 213–231. doi: 10.1080/14616734.2012.672262
- Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assess. Eval. High. Educ.* 34, 159–179. doi: 10.1080/02602930801956059
- Sadler, D. R. (2014). The futility of attempting to codify academic achievement standards. *High. Educ.* 67, 273–288. doi: 10.1007/s10734-013-9649-1
- Scheff, T. J. (1990). *Microsociology. Discourse, Emotion and Social Structure*. Chicago: The University of Chicago Press.
- Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. Aldershot: Ashgate.
- Skibsted, E., and Matthiesen, N. (2016). *Relationskompetencer i læreruddannelse og skole [Relational Competence in Teacher Education and School]*. Fredrikshavn: Dafolo.
- Vaismoradi, M., Turunen, H., and Bondas, T. (2013). Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. *Nurs. Health Sci.* 15, 398–405. doi: 10.1111/nhs.12048
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. Cambridge: Cambridge University Press.
- Wubbels, T., and Brekelmans, M. (2005). Two decades of research on teacher-student relationships in class. *Int. J. Edu. Res.* 43, 6–24. doi: 10.1016/j.ijer.2006.03.003
- Wubbels, T., den Brok, P., van Tartwijk, J., and Levy, J. (2012). *Interpersonal Relationships in Education: An overview of Contemporary Research*. Rotterdam: Sense Publishers.
- Yeh, Y.-C. (2004). Nurturing reflective teaching during critical-thinking instruction in a computer simulation program. *Comput. Educ.* 42, 181–194. doi: 10.1016/S0360-1315(03)00071-X
- Yerrick, R., Ross, D., and Molebash, P. (2005). Too close for comfort: real-time science teaching reflections via digital video editing. *J. Sci. Teacher Educ.* 16, 351–375. doi: 10.1007/s10972-005-1105-3

**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 Holmstedt, Jönsson and Aspelin. 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.

## APPENDIX

### Criteria for Analyzing Teachers' Relational Competency

#### Communication

C:1 Teacher's verbal communication is attuned to the student; the teacher focuses on being understood by, and understand, the student.

C:2 The teacher uses verbal and/or non-verbal communication to invite the students to take part in discussions.

C:3 The teacher's non-verbal communication is attuned to the student; the teacher confirms the student through the communication (gestures, ways of speaking, body position, facial expression, etc.).

#### Differentiation

D:1 The teacher maintains an appropriate distance between herself/himself and the student; the teacher is not too far away or too close in her/his relationship with the student.

#### Emotions

E:1 The teacher is sensitive to the student's feelings; the teacher "reads" the student's emotional expressions, responds appropriately, and manages own feelings.

E:2 The teacher acts in order to create a good atmosphere in the group.

#### Professionalism

P:1 The teacher acts responsibly in relationships; she/he appears as can be expected by a professional.

P:2 The teachers meets every student as an individual.



# Are Assessment Exemplars Perceived to Support Self-Regulated Learning in Teacher Education?

Peter R. Grainger\*, Deborah Heck† and Michael D. Carey†

School of Education, Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Maroochydore, QLD, Australia

## OPEN ACCESS

### Edited by:

Anders Jönsson,  
Kristianstad University, Sweden

### Reviewed by:

Graham Hendry,  
University of Sydney, Australia  
David Newlyn,  
Western Sydney University, Australia

### \*Correspondence:

Peter R. Grainger  
peter.grainger@usc.edu.au  
orcid.org/0000-0001-8214-2595

†Deborah Heck  
orcid.org/0000-0002-0235-8546

Michael D. Carey  
orcid.org/0000-0002-3117-9010

### Specialty section:

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

**Received:** 28 March 2018

**Accepted:** 03 July 2018

**Published:** 14 August 2018

### Citation:

Grainger PR, Heck D and Carey MD  
(2018) Are Assessment Exemplars  
Perceived to Support Self-Regulated  
Learning in Teacher Education?  
Front. Educ. 3:60.  
doi: 10.3389/feduc.2018.00060

Assessment exemplars are a tool to guide students to what is valued by assessors in a specific assessment task, in short, as examples which illustrate, typically, dimensions of quality. Often high-quality exemplars are provided in formative assessment contexts to develop self-regulated learning. We were interested in researching the perceived efficacy and impact of a variety of assessment exemplars, ranging from low to high quality, in teacher education courses at a regional university. More specifically, this research explores student perceptions of how assessment exemplars support the development of phases and signposts for self-regulated learning. We surveyed 72 students and found that students accessed exemplars regularly and found them useful in providing detailed guidance that went beyond the descriptions of assessment tasks found in course outlines and assessment rubrics. They valued various types of exemplars, a range of quality, and the inclusion of annotated and unannotated versions of exemplars. We identified four key themes from the analysis: assessment exemplars as guides, supplements, starting points, and standards for comparison. Our results support the provision of exemplars as a tool to build student self-regulation in three phases and their contribution to the four signposts on the path from social to independent self-regulatory practice (Zimmerman and Kitsantas, 2014).

**Keywords:** assessment, exemplars, pre- service education teachers, efficacy, feedforward

## INTRODUCTION

Assessment exemplars are used in a variety of educational contexts (e.g., law, nursing, education) as a formative tool to guide students to what is valued by assessors in a specific assessment task, in short, as examples that illustrate, typically, dimensions of quality. Exemplars are used by students and teachers to develop student self-monitoring and/or self-regulation, to build student self-efficacy and to encourage ownership over learning (Hawe et al., 2017). The aim of the development of these self-regulatory practices is to improve academic performance. Research across several decades suggests a strong link between self-regulated learning and academic achievement (Panadero, 2017). Exemplars can be used as means of experiential learning in which the participants experience exemplars as learners, gaining understanding of the benefits and pitfalls and consequently applying this knowledge in future contexts upon graduation (Dixon and Hawe, 2016). In the context of teacher education, developing, and experiencing the impact of self-regulatory practices on learning provides an important contribution to both preservice and in-service teachers' developing professional practice (Panadero, 2017).



The word exemplar is defined as “key examples chosen so as to be typical of designated levels of quality or competence” (Sadler, 1987, p. 200). Carless and Chan (2017, p. 1) define exemplars “as carefully chosen samples of student work which are used to illustrate dimensions of quality and clarify assessment expectations.” Newlyn (2013) describes them as examples of best or worst practice designed to promote student understanding of particular skills, content, or knowledge in addition to their use in articulating criteria and standards for assessment tasks. In some cases, these exemplars are provided from a pool of assessment work that has been produced by a previous cohort of students. Typically, these are high quality examples of student work. It is also the case, although less commonly practiced, that assessment exemplars illustrating poor quality can also be provided as guides to students.

Sadler (2002) noted that exemplars convey messages about quality or lack of quality that no other mechanism can provide. They act as a performance benchmark for students by which their own performance can be evaluated and honed. Exemplars offer an “embodiment of standards” (Sadler, 2005, p. 190). Similarly, Bell et al. (2013) defined exemplars as “illustrations of assessment standards in practice” (p. 771). Hence, they also serve not only to improve student outcomes on a task, but they also act as a self-evaluation tool that encourages students to make their own informed judgments (Carless, 2015) about the nature of quality. Similarly, Scoles et al. (2013), Hawe et al. (2017), and Carter et al. (2018) refer to the self-regulatory nature of exemplars as a “feedforward” mechanism supporting students when writing academically. These researchers also noted the impact of exemplars on motivation, self-efficacy, and self-monitoring, in addition to their positive impact on understanding task requirements and the structure of academic tasks, support, and advancement of subject knowledge.

To improve student outcomes in assessment tasks, assessors also commonly provide written feedback in the form of annotations on student work. This feedback mechanism has been discussed in the assessment literature at length, a major finding being that student feedback is not well-understood by students, or ignored, and a constant source of frustration for both assessor and student alike (Grainger, 2015). Exemplars provide a clarity (Price et al., 2012) that rubrics are often criticized for not providing, due to fuzziness, or vagueness. In this regard, researchers (Handley and Williams, 2011; Hendry, 2013) report that the use of exemplars complements traditional written feedback mechanisms as students are able to decode the written feedback as a result of their engagement with the exemplars. Recently, To and Liu (2017) researched the impact of peer and teacher-student exemplar dialogues to unpack assessment standards.

The positive use of exemplars is tempered by the possibility that students may feel that the exemplar provided is the only way to a good result, and hence may in fact restrict creativity, and may result in plagiarism (Newlyn, 2013; Thomson, 2013). Some research into exemplars suggests that students might look at the exemplars alone, without linking them to criteria and descriptors, use them as templates, or plagiarize them (Bell et al., 2013). In their study, Bell, Mladenovic and Price reported 11 students who

found the resources, particularly the annotated exemplars, to be too detailed and prescriptive.

Against this background, we were interested in exploring the value of assessment exemplars at the tertiary level and from a student perspective. We were interested in knowing if exemplars provided online would be accessed by students, how often they were used, in what manner they were used and if students valued a range of exemplars reflecting various standards of quality. In addition, we were interested to know how students perceived the provision of exemplars that exemplified a FAIL standard and how these were valued in comparison to those that exemplified PASS to HIGH DISTINCTION standards. Additionally, we were interested in student perceptions of the value of exemplars as compared to the value of explicit criteria used in rubrics and whether the exemplars we provided supported the assessment task rubrics. Hence, the research questions were:

- Do students value exemplars?
- How often are they accessed by students?
- How do assessment exemplars assist students in determining quality?
- What kind of exemplars assist? (i.e., annotated exemplars or un-annotated exemplars)
- Do exemplars from all achievement standards support student learning?

## LITERATURE REVIEW

As early as 1987, Sadler advocated the use of exemplars to illuminate what he calls “fuzzy standards” (p. 202), and more recently he noted that “[t]he number of exemplars can probably be made fairly small provided they are accompanied by explicit annotations of the properties of individual pieces” (Sadler, 2009, p. 207). Despite this early identification, it is in recent times that the research into the use of exemplars has gained some momentum and largely as a result of the failure of traditional models of feedback in improving student grades (Price et al., 2011). According to Hendry (2013) one-way after-task feedback is not effective. In a similar vein, Scoles et al. (2013) and Wimshurst and Manning (2013) proposed the feedback emphasis be moved to “feedforward” through the provision of exemplars when introducing tasks. This proposal was supported by their quantitative findings which found students who accessed exemplars scored better than those who did not.

Despite this shift toward the use of exemplars, a search of the assessment literature failed to reveal a single study that dealt with the use of online assessment exemplars at the tertiary level in preservice teacher education courses. Hence, we conclude that there is a gap in the assessment literature that is addressed by our current study and the need to address this gap is echoed by Bell et al. (2013, p. 771) who noted “Little is known about how students use these resources although they are regarded as positive.”

Specific to our context of providing online exemplars at tertiary level, we note the work of Handley and Williams (2011) who reported that in a cohort of 400 students most students were receptive to online annotated exemplars which they found

to be very useful in terms of providing guidance on structure and layout and clarifying expectations. In fact, some found it motivating, in that they reportedly wanted to match or beat the quality standard. Some students wanted examples of poor assignments, a result which is also investigated as a focus of our study.

Although some studies (Rust et al., 2003) have reported improved student outcomes as a result of using exemplars, other studies (Carter et al., 2018) have found the benefits of exemplars were not reflected in improved student performance, despite their perceived efficacy by students. Other studies (Newlyn et al., 2012) found no significant impact, positive or negative over a 4-year period. We also note, specific to our context, the study by Hendry et al. (2011) on the use of a variety of exemplars in a first-year law course, reflecting different standards (poor, borderline, and excellent). They reported the usefulness of the templates to students in providing direction and ideas, and to assessors to explain standards through exemplars. In their study, they found evidence that when students engage in the task of marking exemplars, accompanied by teacher explanations of the grades awarded to the exemplars, students develop a better understanding of quality and hence assessor expectations (Hendry et al., 2011). This finding was replicated in further studies by Hendry and Anderson (2012) and Hendry and Jukic (2014) which also reported that students valued interactive assessor explanations for grades given to exemplars.

Similarly, Kean (2012) identified peer assessment, not just interaction and marking by an instructor, alongside the provision of exemplars, as a formative strategy for developing student understanding of quality. Formative assessment is a strategy that can be used to empower students as self-regulated learners (Nicol and Macfarlane-Dick, 2006) reducing a dependency that positions students as passive subjects (Boud, 2007). Panadero et al. (2018, p. 13) suggest that “self-regulated learning should be the primary goal of formative assessment.” Exemplars are not standards but are indicative of standards and when accompanied by marking guides and dialogue and/or student marking of exemplars, they can have a significant impact on learning as students learn to interpret and apply standards to recognize “quality.” In this regard, the literature on self-regulation is lengthy. For example, Perry and Smart (2007) define it as “...an active constructive process whereby learners set goals for their learning and monitor, regulate, and control their cognition, motivation, and behavior, guided, and constrained by their goals and the contextual features of the environment” (p. 64). While most research is focused on the development of specific interventions that support learning in the three common phases evident in self-regulated learning models (preparation, performing, and reflection), it is acknowledged that little research focusses on assessment specifically (Panadero et al., 2018).

In this study we draw upon the socio cognitive work of Zimmerman (2000, 2002) and Zimmerman and Kitsantas (2014). They suggest a three-phase theoretical model of self-regulation, which refers to “self-generated thoughts, feelings, and behaviors that are oriented to attaining goals” (p. 65). The three phases

include a forethought phase (processes and beliefs that occur before efforts to learn), a performance phase (processes that occur during implementation), and a self-reflection phase (processes that occur after each learning effort). Zimmerman and Kitsantas (2014) suggest that there are four signposts that support the move toward that attainment of self-regulatory competence. The first sign post is at an *observation* level where learners explore model performances to support their understanding of the skill or task. The second signpost is *emulation* where the learner uses the model response to generate their own version of the task. At the third *self-controlled* signpost the learners apply the ideas to writing beyond the scope of the examples provided, and at the final signpost the *self-regulated* level allows the learner to practice and make adaptations based on their own experiences.

Assessment exemplars provide opportunities for students and teachers to explore assessment task requirements. Most recently, To and Carless (2015) and To and Liu (2017) reported on studies that focussed on the “dialogic use of exemplars” which they described as student participation in discussion to maximize the potential of analyzing exemplars. They also noted the importance for the teacher that “teacher guidance serves to explicate the characteristics of good quality work and to increase students’ critical awareness of the differences between exemplars and their own writing” (p. 1). Similarly, Carless and Chan (2017) investigated the role of dialogue in supporting students to develop their appreciation of quality work through the use of exemplars. The purpose of this work is to further explore the contribution of assessment exemplars in the context of self-regulation.

To summarize, the research into the efficacy of exemplars tells us that students appreciate the provision of exemplars because exemplars reveal tacit knowledge (To and Carless, 2015) and illustrate assessor perceptions about what is valued in student work. Exemplars can complement traditional written feedback processes; they can be annotated; and provide a stimulus for dialogue and discussion. In addition, they support the transmission of tacit knowledge pertinent to a discipline and they assist in transmitting knowledge of criteria and standards (Newlyn and Spencer, 2009). This paper explores student perceptions of how assessment exemplars support the development of phases and signposts for self-regulated learning.

## METHODS

In this exploratory study we targeted courses at undergraduate and postgraduate level in teacher education programs (undergraduate Bachelor of Education; postgraduate Diploma of Education, postgraduate Master of Education) at a regional university. In total we targeted six courses, with total enrolments of ~300 students. We used the Learning Management System, BlackBoard to upload a variety of assessment exemplars representing a variety of standards of student work from previous iterations of these courses. The exemplars ranged from FAIL to HIGH DISTINCTION. The exemplars were all written essays ranging in length from 2000 to 5000 words. The exemplars were loaded prior to the commencement of the

course. We “enabled statistics” on the BlackBoard sites in order to access usage information. To enable us to answer the research questions, we implemented a student survey consisting of Likert scale and open response questions. Participation was voluntary. Seventy-one percent of students were undergraduate students; 26% were Graduate Diploma students and the remaining 3% were from Master of Education courses.

Data were collected over the course of one semester. Ethical consent was received from the Human Research Ethics Committee at the university in which this study was conducted (protocol number A/16/789) and the consent of the participants was obtained by virtue of survey completion after potential participants were provided with all relevant information. After the data were collected, we undertook a process of first-cycle descriptive coding and second-cycle pattern coding in accordance with Miles et al. (2014). Two researchers independently coded the participants’ responses to the questions in a cross-case analysis to identify themes, similarities, and differences (Creswell, 2007), and then cross-checked for consistency in theme identification before constructing a descriptive narrative to present, analyze, and discuss the findings.

In total we received 72 responses to the survey. As this study was explorative, we focused our study broadly across the teacher education cohort within our education courses, collecting exemplars varying in quality, and with or without annotations (Table 1). Our sampling strategy was therefore convenience sampling, rather than purposive sampling (Rapley, 2014).

We wanted to evaluate the efficacy of a range of exemplars, annotated and unannotated. Our annotations were not annotations in relation to the genre of writing but rather, were explicitly connected to the criteria and standards in the rubric that were used to make judgments about the quality of student work. The mark ups we provided in the form of annotations, identified instances where the evidence (i.e., the student work) exemplified the criteria and standards. If all exemplars were to be annotated, then there would be no way of evaluating if students valued exemplars that were not annotated. Conversely, if all exemplars were to be unannotated, we would not be able

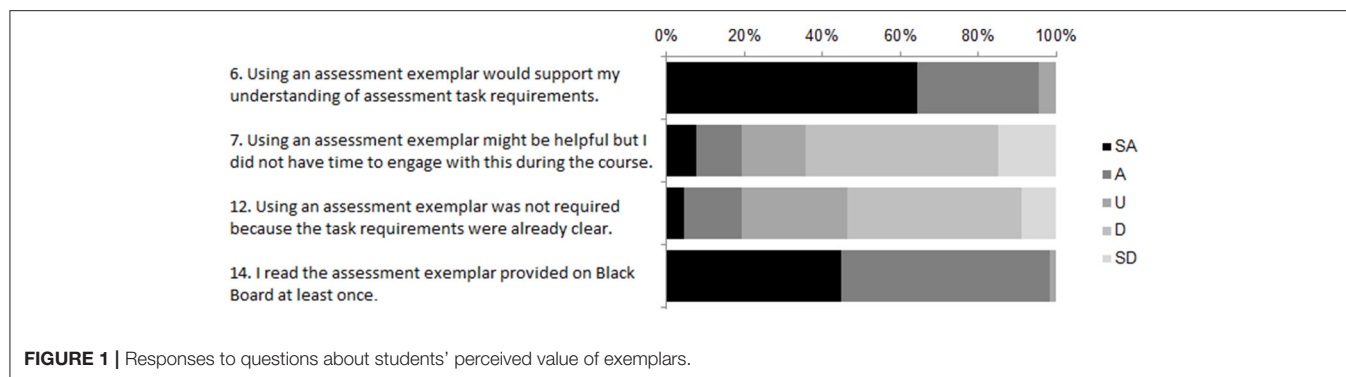
to evaluate directly if students also valued annotated exemplars. We would have to be relying on their comments and there would be no way of ensuring their comments. In some courses we provided the full range of exemplars covering all the standards, but in others we provided single exemplars. In the Masters courses the exemplars were discussed with a tutor and the grades allocated scrutinized against the criteria sheets (also referred to as rubrics) to determine the alignment. In the undergraduate courses and the Graduate Diploma courses, the exemplars were not discussed in any way. We had anticipated a larger number of responses but the small number we received ( $n = 72$ ) meant that we did not have the statistical power to find significance, so our analyses described below, were limited to a narrative based on descriptive statistics. We used a tool known as *page skip logic* to enable students to progress to relevant sections of the survey, depending upon their context.

## Descriptive Statistics

Our first research question was to what degree students valued exemplars; this question was measured by many of the Likert scale questions and a further indication of value was observed through the Blackboard statistics log of frequency with which students accessed the exemplars provided to them. For example, 95% of students surveyed supported our assumption that students would value exemplars due to the support that they would provide students to understand the assessment requirements (Q6). This was confirmed by the statistic that 100% of surveyed students accessed the exemplars provided (Q13) despite 20% of the surveyed students acknowledging the fact that this was not required due to the explicitness of the task requirements (Q12). Surprisingly, one in five students admitted that they did not have time to access the exemplars (Q7) despite the perceived value of doing so. Ninety eight percent of students accessed the exemplars at least once (Q14); 77% accessed the exemplars between two and four times; 17% between five and 10 times and 6% more than 11 times (Q15). The results pertaining to students’ perceived value of exemplars are summarized in Figure 1. The abbreviations to the right of each table refer to the Likert Scale. SA equates to Strongly Agree; A equates to Agree;

**TABLE 1** | Summary of exemplars provided.

| Program of study      | High distinction | Distinction | Credit | Pass | Fail | Annotated  |
|-----------------------|------------------|-------------|--------|------|------|------------|
| Graduate diploma      | X                |             |        |      | X    | YES        |
| Undergraduate primary | X                | X           | X      | X    | X    | YES and NO |
| Undergraduate primary |                  |             |        | X    |      | NO         |
| Postgraduate masters  |                  |             |        | X    |      | YES        |
| Postgraduate masters  |                  |             |        | X    |      | NO         |
| Postgraduate masters  |                  |             |        | X    |      | YES        |



U means Unsure; D means Disagree; and SD means Strongly Disagree.

We were especially interested in what kind of explicit support the exemplars provided to students (**Figure 2**). Ninety-five percent indicated that the exemplars clarified *expectations* (Q19); 63% of students valued the support in terms of *content* (Q18); and 54% valued the support in terms of understanding *academic literacies* (Q22).

In addition, we were interested to know how many and what kinds of exemplars were most valued by students (**Figure 3**), that is, would students value the utility of HIGH DISTINCTION exemplars as well as FAIL exemplars. In this regard, 88% of students surveyed (Q8) valued the provision of just two exemplars at both ends of the quality continuum (HIGH DISTINCTION and PASS) but just over half of the students (55%) wanted to see an exemplar for every standard (Q9). As we suspected, based on previous literature that we accessed, 43% of students were tentative about using exemplars due to possible plagiarism complications (Q11).

We wanted to know if students wanted in-class discussions of the exemplars—clean, un-annotated copies of exemplars, or annotated exemplars—and how these supported their understanding of the assessment task. To determine students' attitudes regarding the utility of *unpacking* the exemplars via “discussion,” we provided two questions: (Q21) 72% valued the in class discussions of the exemplars provided, and (Q10) 33% valued these discussions via an online discussion board. Annotated or not, students found the exemplars useful, evidenced by their responses to Q16 (70%) and Q17 (82%) despite the fact that only 58% understood the annotations and comments provided on the exemplar (Q20). These responses are identified in **Figure 4**.

Finally, some questions (Q7, Q9, Q10, Q12, Q16, Q18, Q20, Q21, Q22) triggered “unsure” responses, identified by us, the researchers, as being at least 10% of the responses, particularly for those questions that involved the use of annotations.

## Thematic Analysis

To validate and verify the quantitative responses reported above we also included in our survey an open response section, consisting of three questions.

Q24: What specifically was beneficial about having exemplars?

Q25: Do you think you will get a better grade as a result of having exemplars provided for you?

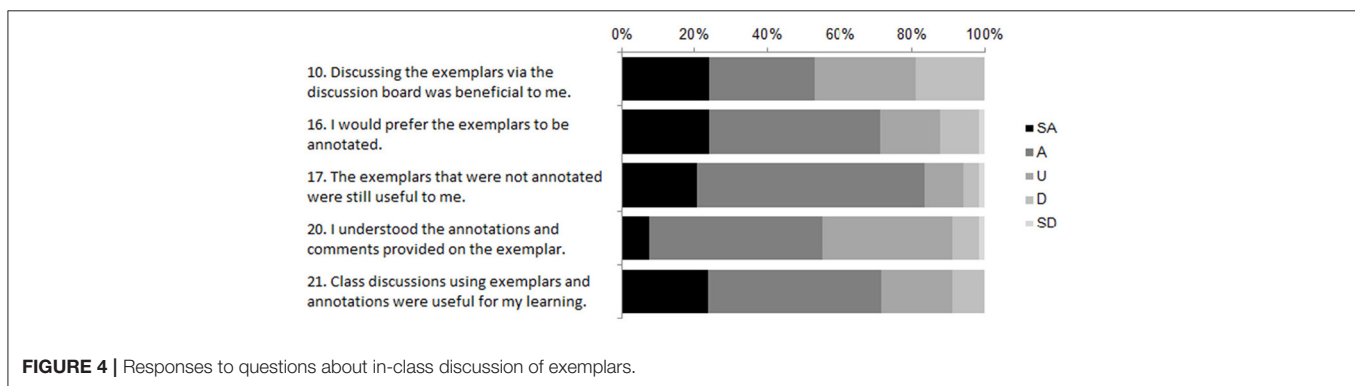
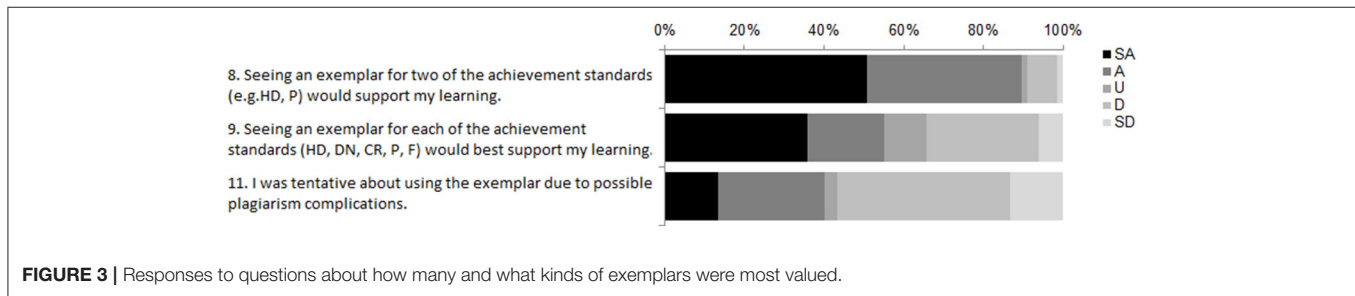
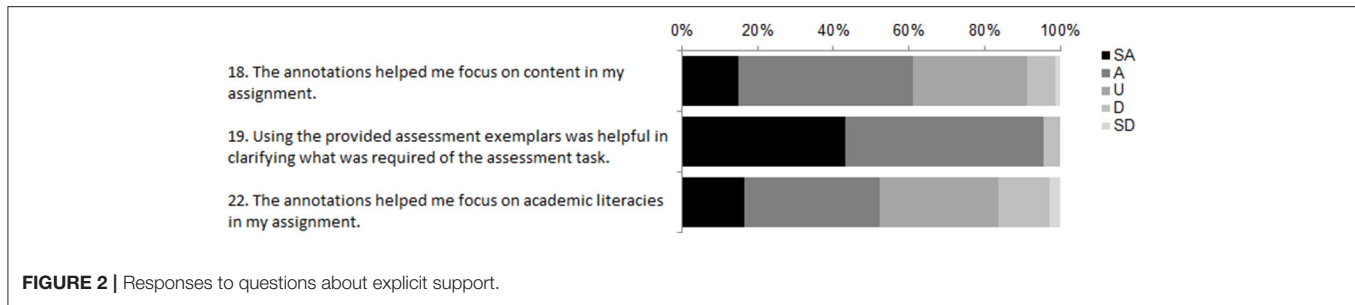
Q26: Please provide one example of how the exemplar improved your assessment task? Why exactly?

Students valued the exemplars because they provided a clarification of the standards expected by the assessor, both specifically and also as a general guide. The exemplars provided not just a starting point but an end point, a target or even a benchmark to be met, and a way to compare their own scripts with various standards. They were valued not just in their own right, but also as a complement to the marking guides, and criteria sheets provided. More specifically, the exemplars provided a clear expectation of desired structures, sequences, and layout, which were not necessarily included in the task guidelines, due to the limitations of space. In particular, the exemplars provided tangible examples of the expectations of academic literacies, including referencing. The final theme that emerged was unexpected in terms of devaluing the FAIL exemplar as unhelpful, demotivating and even distracting. We believed that this exemplar would be valued by students, but this was not the case. Hence, four key themes form the findings from the analysis: assessment exemplars as guides, supplements, starting points, and standards for comparison.

Our first theme identified the exemplars as useful to students because they provided a guide, both generally and explicitly. One student referred to the distinction level exemplar as a “bible” that would be followed to achieve this particular standard. While others remarked on the usefulness of the exemplar in providing “*clear directions as to what the task expectations were*” in particular in regard to “*each of the standards*” illustrating the difference between a pass and a high distinction standard for example. The exemplars also provided specific guidance in terms of layout, content, language to be used, format and structure of the assignment. One student said “*The exemplars gave me guidance as to the structural layout of the assignment. I found it very useful to see how the breakdown of paragraphs and sections could be used to add emphasis.*” Another commented “*Exemplars allowed me to see the way the assessment was set out and being able to see the structure of the writing.*”

Our second theme is related to the usefulness of exemplars as “*valued supplements to the criteria sheet and task guidelines*”. Students commented that criteria sheets/rubrics “*were not always*





clear” and were “quite subjective,” even “confusing.” Exemplars provided a “clarity” that the rubric and the task description alone could not achieve. One student commented “You get an opportunity to see mistakes or what could have been improved to get a higher mark that is not explicit in the criteria, and then adjust your assessment piece accordingly.”

Expectations of academic literacies and referencing requirements were a major theme, typified by comments such as “It made the expected writing quality clear” and it provided “clarity on the topic and on academic literacy.” We were surprised by the fine-grained nature of the engagement of some students, characterized by the following comments: “It assists in understanding of structure, grammar, punctuation.” And “It was helpful to have a reference list.”

Theme three referred to the usefulness of exemplars as a starting point for the students and in one case, even a motivating factor. “They give an initial spark or direction for me. After that I make the assignment my own.” Students also suggested

that in the context of their busy workload across a variety of tasks and courses the exemplars provided a way into the task because “Sometimes starting is the hardest.” Getting students to begin the process of thinking about their work is an important way to engage the learner, indicated by comments such as the examples “help me start thinking about the task, see how it can come together.”

Theme four identifies exemplars as being useful in terms of how they exemplified a certain standard of work, or a “source of comparison with their own work” and also with the rubric used to assess the task. In short, after the assignment had been done, students accessed the exemplars to “evaluate their own direction,” perform a “final checklist,” a “benchmark” exercise, which was then used to hone their own work. Connected to this final theme of comparison was the student suggestion that providing a FAIL exemplar was not useful. Typically, students found these “not useful,” because they “did not assist learning,” were “demotivating and distracting.”

## DISCUSSION

This study fills a gap in the literature about the perceived efficacy of using online assessment exemplars which according to Handley and Williams (2011), is relatively modest. Our literature review revealed a dearth of literature that discussed the provision of online assessment exemplars at the tertiary level in teacher education courses. It also suggests the importance of teachers engaging in and experiencing self-regulatory learning so that they can develop their own professional practice (Panadero, 2017). Hence, this discussion will explore the contribution of assessment exemplars to Zimmerman's (2002) three phases of self-regulation and the contribution to the four signposts on the path from social to independent self-regulatory practice (Zimmerman and Kitsantas, 2014).

At the forethought phase of self-regulation (Zimmerman, 2002), the assessment exemplars provided students with information that support them in their analysis of the task. Our study reflected the existing literature that students do value exemplars and access them often. The themes of exemplars as guides, supplements and starting points connect with the notion of both task analysis and motivation to get started. In this context the assessment exemplars provided the opportunity to achieve the first signpost in the self-regulation journey, an opportunity to observe an example of the task requirements. However, it must be noted that the provision of the FAIL standard was identified by students as counter-productive and not motivating, a perception not reported before in the literature and contrary to what Handley and Williams (2011) reported as being desired by students. It suggests that students draw upon the exemplar to examine and observe what to do for the purpose of moving toward the stage of emulating the task in accordance with the self-regulatory signposts of Zimmerman and Kitsantas (2014). This suggests that the students are operating with the use of exemplars at the first and second stages, both of which requires social interaction and discussion with their teachers and lecturers to further develop as self-directed learners. Our study supports very strongly, the value of dialogic use of exemplars, in other words, discussions among students and assessors about the standards exemplified in the exemplars. This supports the focus by Hendry et al. (2011), Hendry and Anderson (2012), Hendry and Jukic (2014), and Hendry (2013) who also reported the high value placed on interactive assessor explanations by students.

During the performance phase and the self-reflection phase of self-regulation (Zimmerman, 2002), as students began the work of developing their task, the exemplars provided a standard for comparison. Our study supports previous work by Sadler (2005) in that exemplars provide a benchmark and represent an embodiment of standards or illustrations of standards in practice (Bell et al., 2013). Similarly, our study supports the results by Price et al. (2012) that assessment exemplars provide a clarity that rubrics cannot provide. Hence our study, regardless of the course being studied, reinforces these previous findings. In this regard we note the increasing necessity to provide exemplars for students as support for rubrics that have been typically criticized by students as being unclear and fuzzy. However, despite existing literature (Sadler, 2009) that suggests annotated exemplars are

preferred, many of our students reported a perceived efficacy with or without annotations. This suggests that students are moving toward the third (self-control) and fourth (self-regulation) signposts of Zimmerman's and Kitsantas' (2014) capability to self-regulate.

While our study reflected the existing literature that students do value exemplars and access them often, at the same time we report conflicting evidence that despite the perceived value and the fact that 100% of students accessed the exemplars, 20% of the students could not find time to engage with them and 20% did not find them useful. The 20% response of not having time to use the exemplars aligns with the type of degree being studied by many of these teacher education students, the Graduate Diploma of Education course, which is an intensive two-semester course, as opposed to the undergraduate students, who study for 4 years or eight semesters. Hence, we conclude that the intensive nature of the course being studied is a variable that impacts upon the engagement of students with exemplars, a finding we had not seen in previous literature.

In addition to the intensive nature of the course as a reason why 20% of students did not engage with the exemplars, we interpret this result as a direct consequence of the development of autonomy and self-direction that increases with student experience of academia. In short, we conclude that postgraduate Masters degree students do not need as much support as undergraduate students and hence, exemplars become devalued as students gain regulatory and self-monitoring skills as they progress from undergraduate to postgraduate academic studies. In this regard, we found no existing literature that described this trend other than those studies, described earlier, that reported the usefulness of exemplars as being related to the development of regulatory skills.

While the remaining 20% who did not find the exemplars useful may well be students who are already operating in a self-directed manner and do not have the need to engage with these examples to define the task for them, another possibility is concern about breaching academic integrity. Academic integrity is a concern for many students due to a fear of plagiarism as a direct result of using exemplars as templates to copy (Thomson, 2013; To and Carless, 2015) and hence, this may be a deterrent to more widespread take-up of exemplars to assist student understanding of assessment tasks. To avoid the fear of plagiarism, a future research focus might analyse the impact of exemplars on student results after discussion of the alternative of using exemplars on a different question or topic to the one set for students in their current assignment. This could be coupled with a direct focus on the use of text-matching software in order to convince students that the potential use of exemplars resulting in plagiarism can be overcome and the potential benefits outweigh the risks.

## CONCLUSIONS

Our results support the development of self-regulated learning in students as a result of direct engagement with exemplars. In our study, students' qualitative comments indicated that

students were self-assessing their own work, thereby taking ownership, actively monitoring and regulating their products by managing different processes as they engaged with the exemplars (Nicol and Macfarlane-Dick, 2006). They were setting goals using the exemplars as benchmarks or targets; they were devising strategies to ensure they avoided plagiarism; they were managing resources (i.e., the various exemplars provided) by determining what exactly they were choosing to focus on, whether it was a FAILED exemplar or a quality exemplar or academic literacies. This is in alignment with Zimmerman's three phases and signposts toward self-regulation (Zimmerman, 2002; Zimmerman and Kitsantas, 2014). It also reflects the work of Perry and Smart (2007) who define the construct of self-regulation as the degree to which students can regulate aspects of their thinking, motivation and behavior during learning.

An additional future research focus is to explore the ways in which assessment exemplars can support rubrics and how rubrics can support assessment exemplars and especially the differences in the perceived value of each of these assessment artifacts. There is a danger that one artifact may not support the other and this may lead to different messages regarding the expectations of the assessment task.

Finally, in conclusion, we suggest that research into the provision of assessment exemplars, using greater numbers of students across different courses and contexts, will provide added clarity to the results of the study reported here. In particular, we suggest future research focuses on those issues around the use of annotations, the efficacy of discussing exemplars through

dialogue, the fear of plagiarism, and the large numbers of students who are still unsure about the usefulness of exemplars in improving their understanding of assessment requirements. Regarding this, we suggest further research into the provision of exemplars in the online mode, either as a support for student-teacher conversations or as stand-alone artifacts.

We suggest that future work analyses the growing independence of students as they progress from academic year to academic year to determine if assessment exemplars are most useful in the early part of a student's academic career and if reliance on exemplars decreases over time. In addition, future focuses may want to analyse how perceived efficacy changes over time in direct proportion to the time spent in academia. Our study suggests that this may indeed be the case. As our study was exploratory, future studies could employ an experimental design consisting of three test group conditions provided with varying methods of exemplar input: with/without teacher discussion; with/without annotation; with/without fail exemplars and a control group which receives only a rubric and explanation of the assessment task. This would facilitate answers to a key question: do students who have access to a pass and high-quality exemplars with or without classroom/teacher discussion perform better than students who do not.

## AUTHOR CONTRIBUTIONS

PG implemented the survey and completed the first drafts. DH assisted in the writing of the drafts and referencing. MC analyzed and interpreted the quantitative data.

## REFERENCES

- Bell, A., Mladenovic, R., and Price, M. (2013). Students' perceptions of the usefulness of marking guides, grade descriptors and annotated exemplars. *Assess. Eval. High. Educ.* 38, 769–788. doi: 10.1080/02602938.2012.714738
- Boud, D. (2007). "Reframing assessment as if learning were important," in *Rethinking Assessment in Higher Education*, eds D. Boud and N. Falchikov (Abingdon: Routledge), 14–26. doi: 10.4324/9780203964309
- Carless, D. (2015). *Analysing Exemplars to Support Students' Understandings of Assessment*. CRADLE seminar, Melbourne, VIC.
- Carless, D., and Chan, K. K. H. (2017). Managing dialogic use of exemplars. *Assess. Eval. High. Educ.* 42, 930–941. doi: 10.1080/02602938.2016.1211246
- Carter, R., Salamonson, Y., Ramjan, L., and Halcomb, E. (2018). Students use of exemplars to support academic writing in higher education: an integrative review. *Nurse Educ. Today* 65, 87–93. doi: 10.1016/j.nedt.2018.02.038
- Creswell, J. W. (2007). *Qualitative Enquiry and Research Design: Choosing Among Five Approaches*. London: SAGE Publications.
- Dixon, H., and Hawe, E. (2016). Utilizing an experiential approach to teacher learning about AFL: a consciousness raising opportunity. *Austr. J. Teach. Educ.* 41:1. doi: 10.14221/ajte.2016v41n1.1
- Grainger, P. (2015). How do pre-service teacher education students respond to assessment feedback? *Assess. Eval. High. Educ.* doi: 10.1080/02602938.2015.1096322
- Handley, K., and Williams, L. (2011). From Copying to Learning: using exemplars to engage students with assessment criteria and feedback. *Assess. Eval. High. Educ.* 36, 95–108. doi: 10.1080/02602930903201669
- Hawe, E., Lightfoot, U., and Dixon, H. (2017). First-year students working with exemplars: promoting self-efficacy, self-monitoring and self-regulation. *J. Further High. Educ.* doi: 10.1080/0309877X.2017.1349894. [Epub ahead of print].
- Hendry, G., Bromberger, N., and Armstrong, S. (2011). Constructive guidance and feedback for learning: the usefulness of exemplars, marking sheets and different types of feedback in a first-year law subject. *Assess. Eval. High. Educ.* 36, 1–11. doi: 10.1080/02602930903128904
- Hendry, G. D. (2013). "Integrating feedback with classroom teaching: using exemplars to scaffold learning," in *Reconceptualising Feedback in Higher Education: Developing Dialogue with Students*, eds M. Merry, D. P. Carless, and M. Taras (Abingdon: Routledge), 133–141.
- Hendry, G. D., and Anderson, J. (2012). Helping students understand the standards of work expected in an essay: using exemplars in mathematics pre-service education classes. *Assess. Eval. High. Educ.* 38, 754–768. doi: 10.1080/02602938.2012.703998
- Hendry, G. D., and Jukic, K. (2014). Learning about the quality of work that teachers expect: students' perceptions of exemplar marking versus teacher explanation. *J. Univ. Teach. Learn. Pract.* Available online at: <http://ro.uow.edu.au/jutlp/vol11/iss2/5>
- Kean, J. (2012). Show AND tell: Using peer assessment and exemplars to help students understand quality in assessment. *Pract. Res. High. Educ.* 6, 83–94. Available online at: <https://ojs.cumbria.ac.uk/index.php/prhe/article/view/126>
- Miles, M. B., Huberman, A. M., and Salda-a, J. (2014). *Qualitative Data Analysis: A Methods Sourcebook, 3rd Edn*. Los Angeles, CA: SAGE Publications.
- Newlyn, D. (2013). Providing exemplars in the learning environment: the case for and against. *Univ. J. Educ. Res.* 1, 26–32. doi: 10.13189/ujer.2013.010104
- Newlyn, D., Juriansz, J., and Spencer L. (2012). Using exemplars to improve student performance in a contracts law research essay. *Int. J. Learn. High. Educ.* 19, 83–94.
- Newlyn, D., and Spencer, L. (2009). Using exemplars in an interdisciplinary law unit: listening to the students' voices. *J. Aust. Law Teach. Assoc.* 2, 121–133. Available online at: <file:///usc.internal/usc/home/pgrainge/Downloads/2-2-59-749.pdf>

- Nicol, D., and Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Stud. High. Educ.* 31, 199–218. doi: 10.1080/03075070600572090
- Panadero, E. (2017). A review of self-regulated learning: six models and four directions for research. *Front. Psychol.* 8:422. doi: 10.3389/fpsyg.2017.00422
- Panadero, E., Andrade, H., and Brookhart, S. (2018). Fusing self-regulated learning and formative assessment: a roadmap of where we are, how we got here, and where we are going. *Aust. Educ. Res.* 45, 13–31. doi: 10.1007/s13384-018-0258-y
- Perry, R. P., and Smart, J. C. (eds.). (2007). *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*. Springer.
- Price, M., Handley, K., Miller, J., and O'Donovan, B. (2011). Feedback: all that effort, but what is the effect? *Assess. Eval. High. Educ.* 35, 277–289. doi: 10.1080/02602930903541007
- Price, M., Rust, C., O'Donovan, B., Handley, K., and Bryant, R. (2012). *Assessment Literacy: The Foundation for Improving Student Learning*. Oxford: Oxford Centre for Staff and Learning Development.
- Rapley, T. (2014). "Sampling strategies in qualitative research," in *The SAGE Handbook of Qualitative Data Analysis*, ed U. Flick (London: SAGE Publications Ltd.), 49–63.
- Rust, C., Price, M., and O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assess. Eval. High. Educ.* 28, 147–164. doi: 10.1080/02602930301671
- Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxf. Rev. Educ.* 13, 191–209. doi: 10.1080/0305498870130207
- Sadler, D. R. (2002). "Ah!... So That's Quality," in *Assessment Case Studies, Experience and Practice from Higher Education*, eds P. Schwartz and G. Webb (London: Kogan Page), 130–136.
- Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assess. Eval. High. Educ.* 30, 175–194. doi: 10.1080/0260293042000264262
- Sadler, D. R. (2009). Grade integrity and the representation of academic achievement. *Stud. High. Educ.* 34, 807–826. doi: 10.1080/03075070802706553
- Scoles, J., Huxham, M., and McArthur, J. (2013). No longer exempt from good practice: using exemplars to close the feedback gap for exams. *Assess. Eval. High. Educ.* 38, 631–645. doi: 10.1080/02602938.2012.674485
- Thomson, R. (2013). Implementation of criteria and standards-based assessment: an analysis of first-year learning guides. *High. Educ. Res. Dev.* 32, 272–286. doi: 10.1080/07294360.2012.676026
- To, J., and Carless, D. (2015). Making productive use of exemplars: peer discussion and teacher guidance for positive transfer of strategies. *J. Further High. Educ.* 40, 746–764. doi: 10.1080/0309877X.2015.1014317
- To, J., and Liu, Y. (2017). Using peer and teacher-student exemplar dialogues to unpack assessment standards: challenges and possibilities. *Assess. Eval. High. Educ.* 43, 449–460. doi: 10.1080/02602938.2017.1356907
- Wimshurst, K., and Manning, M. (2013). Feed-forward assessment, exemplars and peer-marking: evidence of efficacy. *Assess. Eval. High. Educ.* 38, 451–465. doi: 10.1080/02602938.2011.646236
- Zimmerman, B. J. (2000). "Attainment of self-regulation: a social cognitive perspective," in *Handbook of Self-regulation*, eds M. Boekaerts, P. R. Pintrich, and M. Zeidner (San Diego, CA: Academic Press), 13–39.
- Zimmerman, B. J. (2002). Becoming a self regulated learner: an overview. *Theor. Pract.* 41, 64–70. doi: 10.1207/s15430421tip4102\_2
- Zimmerman, B. J., and Kitsantas, A. (2014). "The hidden dimensions of personal competence: self-regulated learning and practice," in *Handbook of Competence and Motivation*, eds A. J. Elliot, C. S. Dweck, and M. V. Covington (New York, NY: Guilford Press), 509–526.

**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 Grainger, Heck and Carey. 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.





# Applying Criteria to Examples or Learning by Comparison: Effects on Students' Evaluative Judgment and Performance in Writing

Renske Bouwer<sup>1\*</sup>, Marije Lesterhuis<sup>1</sup>, Pieterjan Bonne<sup>2</sup> and Sven De Maeyer<sup>1</sup>

<sup>1</sup> Training and Educational Sciences, University of Antwerp, Antwerp, Belgium, <sup>2</sup> Artevelde University College Ghent, Ghent, Belgium

## OPEN ACCESS

### Edited by:

Frans Prins,  
Utrecht University, Netherlands

### Reviewed by:

Jill Willis,  
Queensland University of Technology,  
Australia  
Peter Ralph Grainger,  
University of the Sunshine Coast,  
Australia

### \*Correspondence:

Renske Bouwer  
renske.bouwer@vu.nl

### Specialty section:

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

**Received:** 30 April 2018

**Accepted:** 14 September 2018

**Published:** 11 October 2018

### Citation:

Bouwer R, Lesterhuis M, Bonne P and  
De Maeyer S (2018) Applying Criteria  
to Examples or Learning by  
Comparison: Effects on Students'  
Evaluative Judgment and Performance  
in Writing. *Front. Educ.* 3:86.  
doi: 10.3389/feduc.2018.00086

In higher education, writing tasks are often accompanied by criteria indicating key aspects of writing quality. Sometimes, these criteria are also illustrated with examples of varying quality. It is, however, not yet clear how students learn from shared criteria and examples. This research aims to investigate the learning effects of two different instructional approaches: applying criteria to examples and comparative judgment. International business students were instructed to write a five-paragraph essay, preceded by a 30-min peer assessment in which they evaluated the quality of a range of example essays. Half of the students evaluated the quality of the example essays using a list of teacher-designed criteria (criteria condition;  $n = 20$ ), the other group evaluated by pairwise comparisons (comparative judgment condition;  $n = 20$ ). Students were also requested to provide peer feedback. Results show that the instructional approach influenced the kind of aspects students commented on when giving feedback. Students in the comparative judgment condition provided relatively more feedback on higher order aspects such as the content and structure of the text than students in the criteria condition. This was only the case for improvement feedback; for feedback on strengths there were no significant differences. Positive effects of comparative judgment on students' own writing performance were only moderate and non-significant in this small sample. Although the transfer effects were inconclusive, this study nevertheless shows that comparative judgment can be as powerful as applying criteria to examples. Comparative judgement inherently activates students to engage with exemplars at a higher textual level and enables students to evaluate more example essays by comparison than by criteria. Further research is needed on the long-term and indirect effects of comparative judgment, as it might influence students' conceptualization of writing, without directly improving their writing performance.

**Keywords:** criteria, comparative judgment, exemplars, peer assessment, writing, evaluative judgment

## INTRODUCTION

In higher education, writing tasks are often accompanied by rubrics or lists of criteria indicating key aspects of writing quality. The primary aim of these analytic schemes is to support teachers in evaluating the quality of students' writing performance. Sometimes teachers also share the criteria with students before they start writing their text. The wide-held

belief is that when students know what aspects are related to quality performance, that they can apply this knowledge successfully to their own performance. However, it can be questioned whether merely sharing teacher-designed criteria with students has the desired effect on students' learning and performance. According to Sadler (1989, 2002), criteria may well explain how the work will be graded, but they do so in rather discrete and abstract terms (e.g., is this text coherent or not), without revealing how the criteria are visualized in a text and how they interactively contribute to the overall quality of a text. This is especially relevant in the context of learning to write, as text quality is more than the sum of its constituent parts (Sadler, 2009). Even rubrics, which specify the performance levels and standards for each of the criteria, can include descriptions that are too abstract for students to truly understand what writing quality entails (Brookhart, 2018). Therefore, Sadler as well as other prominent scholars in the field of assessment (cf. Boud, 2000; Nicol and Macfarlane Dick, 2006; Carless and Boud, 2018) have argued that the relevance of showing examples to students, as "exemplars convey messages that nothing else can" (Sadler, 2002, p. 136). Through the analysis of examples students can experience themselves how high-quality texts are different from average ones, which increases their tacit knowledge of what constitutes text quality, making criteria and standards concrete (Orsmond et al., 2002; Rust et al., 2003; Handley and Williams, 2011).

However, as with most instructional practices, just providing students with examples is insufficient. They should not be seen as model texts that students can copy, but rather as illustrations for which some kind of analysis is necessary to come to a deep understanding of how different dimensions of quality come together (Sadler, 1989; Handley and Williams, 2011; Carless and Chan, 2017). Recently, Tai and colleagues have argued for precisely this shift in education: instead of students being passive recipients of what is the expected standard in their work, they need to actively engage with criteria and examples of varying quality (Tai et al., 2017). There are, however, different ways for doing so, ranging from analytic discussions of only one or two exemplary texts (cf. Carless and Chan, 2017), to comparing and contrasting a number of examples of varying quality (cf. Sadler, 2009). This leaves us with the question how students ideally engage with examples in order to optimize their learning. The aim of the present study is to experimentally investigate whether the way students engage with examples has an impact on their conceptualization of writing quality as well as on their writing performance.

A promising way to provide students with the opportunity to actively engage with examples of varying quality is through the implementation of peer assessment activities (Carless and Boud, 2018). In a peer assessment, examples are authentic pieces of work created by peers, which are therefore quite comparable to the student's own writing. Theories on formative assessment describe that the ability to make qualitative judgments of a peer's work has an effect on how students monitor and regulate the quality of one's own performance (Sadler, 1989; Tai et al., 2017). Self-monitoring and self-regulation skills appear to be a strong predictor of high-quality performance, especially in the context of writing (Zimmerman and Risemberg, 1997; Boud,

2000). Moreover, when students provide peer feedback they need to diagnose strengths and weaknesses in a text and elaborate on possible solutions through which their peers can move forward. This kind of problem-solving behavior asks for a deep cognitive process, which generally has a stronger effect on students' learning than merely receiving feedback (Nicol et al., 2014). By doing so, peer assessment can be used as a pedagogical strategy, not just for assessment purposes, but also for teaching students the content of a course (Sadler, 2010). It is, however, quite a challenge for students to make a deep cognitive analysis of their peers' work, and to provide qualitative feedback accordingly. Students often perceive the quality of the peer feedback as poor, with comments provided at a too superficial level (Patton, 2012; Yucel et al., 2014). In particular, students have the tendency to focus in their feedback at form rather than at content, and they praise their peers more than teachers do (Patchan et al., 2009; Huisman et al., 2018).

To optimize the learning benefits of peer assessment, teachers should support students in how to address both higher and lower level aspects in their feedback. One way to do so is to let students explicitly link the quality of a peer's work to predefined assessment criteria (Rust et al., 2003; Hendry et al., 2011; Carless and Chan, 2017). Although this instructional practice can be effective for peer assessments, an important remark needs to be made. It is not easy for students to use teacher-designed criteria, especially when they do not yet possess a clear understanding of what text quality looks like (Sadler, 2002, 2009). Hence, merely sharing criteria with students is not deemed sufficient. In addition, students may perceive predefined criteria as demands by the teachers, which is associated with only shallow learning and performance (Torrance, 2007; Bell et al., 2013). More beneficial approaches seem to be interactive teacher-led discussions on how to apply assessment criteria to examples (Rust et al., 2003; Bloxham and Campbell, 2010; Hendry et al., 2011, 2012; Bell et al., 2013; Yucel et al., 2014; To and Carless, 2016; Carless and Chan, 2017), or involving students in the developmental process of criteria-based rubrics (Orsmond et al., 2002; Fraile et al., 2017). Drawbacks of such practices are, however, that an effective implementation demands considerable time and resources from teachers, as well as skills to adequately guide students in the peer discussions (Carless and Chan, 2017). In addition, it can be questioned whether breaking down holistic judgments into more manageable parts supports students in grasping the full complexity of judging multidimensional performances (Sadler, 1989, 2009, 2010).

An alternative approach for engaging students with examples of their peers is through learning by comparison. In this approach students are presented with pairs of texts and for each pair they have to indicate which one out of two is the best. It has been established that, even in the absence of evaluation criteria, the process of comparative judgment is easier and leads to more accurate evaluations of quality than absolute judgments in which products are evaluated one by one (Laming, 2004; Gill and Bramley, 2013). In addition, a recent meta-analysis shows that peers are as reliable in making comparative judgments as expert assessors (Verhavert et al., submitted), and that their judgments largely correspond (Jones and Alcock, 2014; Jones and Wheadon, 2015; Bouwer et al., 2018).

Although comparative judgment is originally designed as a method to support assessors in making qualitative judgments (Pollitt, 2004), there is an increasing number of studies pointing toward its potential learning effects (cf. Bouwer et al., 2018). For example, Gentner et al. (2003) found that undergraduate business school students who compared two negotiation scenarios were over twice as likely to transfer the negotiation strategy to their own practice as were those who analyzed the same two scenarios separately, even without any preceding training. Bartholomew et al. (2018b) demonstrated that design students who were part of a comparative-based peer assessment outperformed students who only shared and discussed their work with each other. In open-ended questionnaires afterwards, these students indicated that they especially liked to receive feedback from more than one or two students and that the procedure allowed them to get inspiration for their own work from seeing a wide variety of examples.

Research also suggests that the process of comparing multiple examples requires critical and active thinking, through which students learn the most important features for a particular task. For instance, Kok et al. (2013) revealed that medical students who compared images showing radiological appearances of diseases with images showing no abnormalities learned to better discriminate relevant, disease-related information than students who only analyzed radiographs of diseases. This resulted in improved performance on a subsequent visual diagnosis test. These learning benefits seem to be especially prominent when examples are of contrasting quality. Lin-Siegler et al. (2015) showed that 6th grade students who were presented with stories of contrasting quality wrote stories of higher quality and were more accurate in identifying aspects in their own text that needed improvement compared to students who were presented with only good examples.

Hence, through the process of comparing concrete examples students gradually develop an abstract schema for quality consisting of features that distinguish good from poor quality, which they can use as a benchmark for comparing and evaluating their own work. Whether the learning effects of comparative judgment are more powerful than those of criteria use is not yet investigated, neither are the potential transfer effects to students' own writing capabilities.

## AIM OF THE PRESENT STUDY

The aim of this study was to compare the learning effects of an analytic approach for the evaluation of essays written by peers to a comparative approach in which students evaluate previous essays by comparison. There were two specific research questions in this study. First, we examined the effects of these instructional approaches on students' evaluative judgments of writing quality. For this research question we investigated the reliability and validity of students' evaluations as well as the content of their peer feedback. Together, this will provide an in-depth insight into the effects of the instructional approach on students' conceptualization and evaluation of writing quality. Second, we examined whether possible effects of the instructional

approach for the peer assessment transfer to the quality of students' own writing. As these effects might be moderated by individual differences between students in their knowledge and self-efficacy for writing, we tested and controlled for these individual characteristics.

## METHODS

### Participants

In an authentic classroom context at a university college in Flanders, Belgium, 41 second year bachelor students in business management were instructed to complete a peer assessment of five-paragraph essays in English (L2) and to write a five-paragraph essay in English themselves for the course International Trade English 2A in class. Both tasks were intended as a learning experience for students, they did not receive grades for any of the tasks. There was one student who did not allow us to use the collected data anonymously for research purposes. The data from this student was removed before proceeding with further analysis. Hence, the final dataset consisted of 40 participants, of which 23 female and 17 male students, with a mean age of 19 years (min = 18, max = 22). Dutch was the native language for the majority of students, with the exception of two students who had French as their native language.

### Materials and Procedure

The procedure of the present study consisted of three consecutive phases. In the first phase students were informed about the general aims of the study, i.e., to get insight into how peer evaluation of essays contributes to one's writing performance. In addition, they were informed that all data would be treated anonymously and used only for research purposes, and that the study results would not impact their grades. After signing the informed consent, students were asked to fill in a questionnaire that included questions about their demographic characteristics, self-efficacy for writing and background knowledge of writing five-paragraph essays.

The self-efficacy for writing scale (Bruning et al., 2013) consisted of 16 items that measured students' self-efficacy for ideation (5 items, e.g., I can think of many ideas for my writing,  $\alpha = 0.70$ ), self-efficacy for conventions (5 items, e.g., I can spell my words correctly,  $\alpha = 0.81$ ) and self-efficacy for the regulation of writing (6 items, e.g., I can focus on my writing for at least 1 h,  $\alpha = 0.74$ ). As individual writing performance varies largely between genres (cf. Bouwer et al., 2015), one question was added to measure self-efficacy for writing in this particular genre (i.e., I can write a five-paragraph essay). All items are measured on a scale ranging from 0 to 100. Positive but moderate correlations between the subscales confirmed that the scales are only weakly related, and hence, measure different dimensions of students' self-efficacy for writing ( $0.25 < r < 0.38$ ,  $p < 0.11$ ). The additional question on self-efficacy for writing a five-paragraph essay had a moderate correlation with the subscale of self-efficacy for conventions ( $r = 0.51$ ,  $p < 0.01$ ), and correlated to a lesser degree with self-efficacy for ideation ( $r = 0.39$ ,  $p < 0.05$ ) and self-efficacy for regulation ( $r = 0.40$ ,  $p < 0.05$ ).

The knowledge part of the questionnaire consisted of 10 open and closed-item questions that measured students' genre knowledge of five-paragraph essays. Before class, students were instructed to study the crucial genre elements of a five-paragraph essay through a slidecast and/or a reader. Knowledge questions were focused on the information in this material and included questions about the crucial elements of the introduction and conclusion in a five-paragraph essay, definitions of topic and subtopic sentences, how to provide support for topic sentences and how to create coherence and unity in the text. In the open-ended questions, students had to indicate what distinguishes a good essay from a weak one (i.e., quality characteristics), and how this genre is different from other types of texts (i.e., genre characteristics).

In the second phase, which lasted for 30 min, students had to peer evaluate five-paragraph essays of last year's students. Ten example essays were available. The topic of the essays was related to doing business abroad, in which students either compared a self-chosen country to Belgium according to the most interesting and relevant cultural dimensions of Hofstede (for more information, see [www.hofstede-insights.com](http://www.hofstede-insights.com)), or they explained why they should (not) export a self-chosen business (field) to a certain country. These topics had been discussed in class in the week before. Consequently, all students had the required domain knowledge for evaluating the content of the essays. According to the formal requirements for this writing prompt, essays were within one page (Calibri 11, interspace 1), and references to sources were in accordance to APA norms. The selection of the essays for this peer evaluation was based on the grades for the essays received in the previous year, in such a way that the essays represented the full range of quality from (very) low, over average to (very) high quality.

For the peer evaluation, students were randomly divided into two conditions. Half of the students ( $n = 20$ ) received a criteria list to evaluate the essays analytically, the other students ( $n = 20$ ) evaluated the essays holistically through pairwise comparisons. Students in the criteria condition were instructed to login to Qualtrics, an online survey platform, in which essays were presented to the students in a random order. Students had to read and evaluate each essay one by one on the computer screen, using the following four sets of criteria: (1) content and structure: does the essay include all required elements of a title, introduction, body, and conclusion, the visual and logical structure of the text, and relevance of content for business students, (2) grammatical accuracy: whether the essay is free from grammatical and spelling errors or inconsistencies, and includes fluent sentences, (3) coherence: whether the essay includes linking words, paraphrases, support, and the content shows unity with only one topic per paragraph and a central overall topic, and (4) vocabulary: whether the essay shows a good range of vocabulary that is related to topic, and is formal, specific and varied. For each of these four criteria, students had to provide a score between 0 (not good at all) and 6 (very good). The evaluation grid describing the criteria is provided in **Appendix A**.

In the comparative judgment (CJ) condition students were instructed to login to D-PAC, Digital Platform for the Assessment of Competences (2018, Version 0.13.6), in which they were online presented with pairs of essays. See **Figure 1A** for a screen capture of comparative judgment in D-PAC. For each pair, students had to individually indicate which essay they think is best regarding its overall quality. To support students in making the holistic comparative judgments, they were provided with the same teacher-designed quality criteria as applied in the criteria condition. The maximum number of pairwise comparisons per student was 20, but students were allowed to work at their own pace. An equal views algorithm randomly assigned essays to pairs in such a way that the likelihood that a particular student is presented with a new essay is maximized. By doing so, after five comparisons a student will have seen all ten essays. Students who managed to complete the total of 20 comparisons will have evaluated the ten essays four times.

Students in both conditions were also requested (but not obliged) to provide feedback in terms of strengths (positive feedback) and weaknesses (negative feedback) for each essay. As in the criteria condition, feedback was incorporated into the flow of comparative judgment. Thus, after each pairwise comparison, students provided positive and negative feedback to each of the two texts. **Figure 1B** shows how the feedback form is presented on the D-PAC platform. A built-in feature of D-PAC is that the feedback for a particular essay is remembered. This means that when a particular essay is evaluated for a second time, the previous feedback will be automatically presented again. Students are allowed to change this feedback or add new comments to it. As the feedback is presented only after each comparative judgment it is very unlikely that the feedback will influence the judgments that are made.

After the peer assessment, the writing phase started. In this third phase, all students received the same writing prompt as last year. Students were free to choose one out of the two provided topics, and received the same formal requirements (e.g., one page, sources according to APA norms). They received up to 90 min to write their essay. Students were not allowed to leave class until they uploaded their essay into the D-PAC platform for further analysis.

## Data Preparation and Analyses

### Peer Feedback Coding

The peer feedback that students provided in the criteria ( $n = 106$ ) and CJ condition ( $n = 369$ ) were combined into one dataset. As each pairwise comparison in the CJ condition consisted of feedback on two texts, we transformed this dataset in such a way that each row included feedback on only one text. Further, in the CJ condition, students were presented with their previously given feedback once they saw the same essay again. In more than half of the cases, students added new feedback to the already formulated feedback, but in 45% of the cases students did not change their feedback. We excluded all the peer feedback that was identical to the previously formulated feedback from further analysis as this might artificially inflate the probability on a particular type of feedback. This resulted in a total of 203 feedback segments for the CJ condition.



All the positive and negative peer feedback was categorized by the first author according to one of the four quality aspects of writing that are specified in the evaluation grid (see also **Appendix A**): structure and content, grammatical control, coherence and unity, and vocabulary. To establish the reliability of this coding procedure, a random selection of 10 percent of the essays was double-coded by the second author. Corrected for chance, there was substantial to almost perfect agreement between the two raters in the coding of the peer feedback according to the four aspects in both the criteria and CJ condition, see **Table 1**. Rater differences in the categorization of feedback segments were discussed and resolved before the first author continued with the coding of all other feedback segments. This collegial discussion led to the addition of a fifth category in which feedback comments were placed that cannot be categorized into any of the four other categories (i.e., miscellaneous category). This fifth category included feedback on, for instance, the font, use of sources, and the use of a picture. The total number of unique feedback points per text was used as a measure of the amount of feedback.

To test the effect of condition on the amount and content of the peer feedback multiple cross-classified multilevel models were performed taking into account possible variance in the amount and content of peer feedback due to students ( $N = 40$ ) and essays ( $N = 10$ ) (Fielding and Goldstein, 2006). In particular, to estimate the number of aspects that were mentioned per feedback segment, a generalized linear cross-classified multilevel model was performed with condition as a fixed effect, and students and essays as random effects. As the number of aspects is count data, following a non-normal distribution, this model was tested by a poisson distribution. In addition, separate binomial logistic cross-classified multilevel models were applied to estimate the fixed effect of condition for the average probability on feedback in each of the five categories (structure/content, grammatical control, coherence/unity, vocabulary, miscellaneous) for both positive

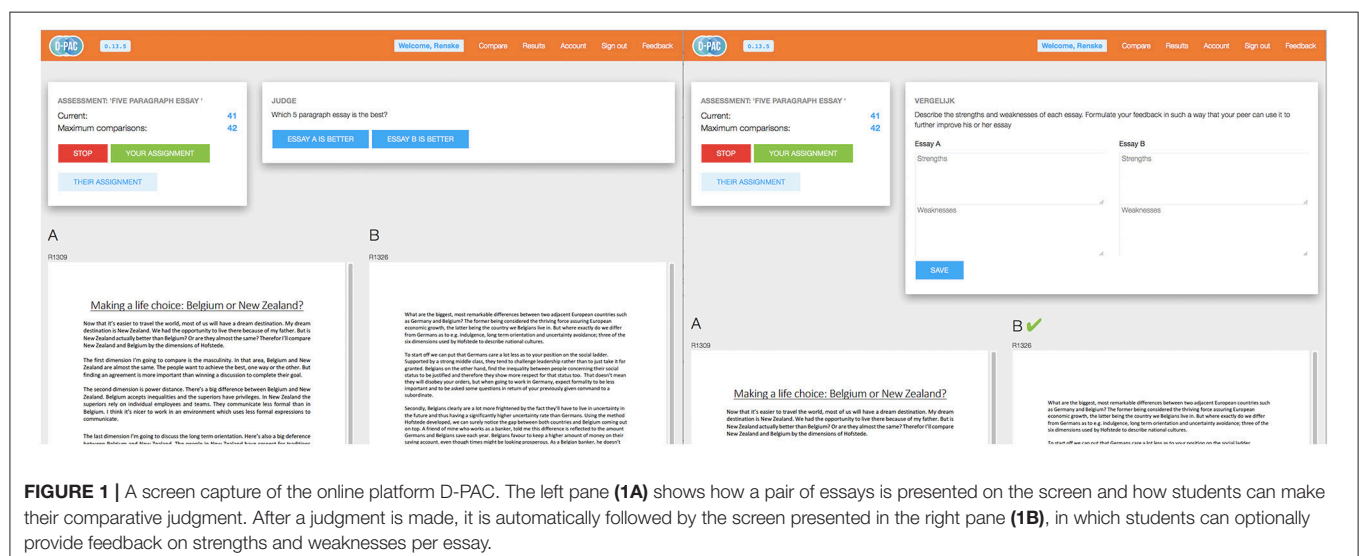
and negative feedback, given a random essay and a random student. The parameter estimates in these models are in logits, which are a nonlinear transformation of the probabilities (cf. Peng et al., 2002). To enhance interpretation the logits are transformed back to probabilities of occurrence.

### Assessment of Essay Quality

The quality of students' own written essays was evaluated by a panel of nine expert assessors using comparative judgment in D-PAC (2018, Version 0.13.6). The panel of assessors consisted of four experienced teachers in business management (three males and one female) and five researchers who are experienced in comparative judgment of writing products (one male and four females). They were instructed to login to the D-PAC platform and complete 40 comparisons in a 4-week period. They were free to do the comparisons when and wherever they wanted. To support the quality of their judgments, assessors were able to consult the students' writing assignment and the assessment criteria at any time. These assessment criteria were the same as the ones that students received during the peer assessment. Of the nine assessors, there was only one teacher and one researcher who did not manage to complete all requested comparisons, they completed only 21 and 24 judgments respectively. Together, the assessors completed 336 comparisons, with each essay being compared 14 to 17 times with a random other.

**TABLE 1 |** Interrater Agreement for Peer Feedback Coding in the Criteria and CJ Condition.

| Aspect                      | Cohen's kappa ( $\kappa$ ) |              |
|-----------------------------|----------------------------|--------------|
|                             | Criteria condition         | CJ condition |
| Structure and content       | 0.75                       | 0.85         |
| Grammatical control         | 1.00                       | 0.93         |
| Coherence and unity         | 0.80                       | 0.76         |
| Formal essay and vocabulary | 0.95                       | 0.98         |



**FIGURE 1 |** A screen capture of the online platform D-PAC. The left pane (1A) shows how a pair of essays is presented on the screen and how students can make their comparative judgment. After a judgment is made, it is automatically followed by the screen presented in the right pane (1B), in which students can optionally provide feedback on strengths and weaknesses per essay.

The Bradley-Terry-Luce model (Bradley and Terry, 1952; Luce, 1959) was used to estimate logit scores for the essays based on the probability that a random assessor assigns a particular essay as the better one, accounting for the quality of the essay to which it is compared. The scale separation reliability of this model was very good, SSR = 0.80, indicating that the estimated logit scores were highly reliable, as were the assessors in their judgments (Verhavert et al., 2017). In addition, there were no individual assessors for whom the pattern of judgments significantly deviated from the estimated model, with standardized likelihood ratios ranging from  $-1.86$  to  $1.22$ .

To estimate the effect of condition on students' writing quality, an independent sample *t*-test with condition (criteria vs. CJ) as the independent variable was performed and the logit scores for writing quality as the dependent variable. As the number of observations in each condition are rather low, we supplement the *p*-values with estimations of effect sizes and confidence to get insight in the magnitude and relative importance of the effects of condition (cf. Nuzzo, 2014; Wasserstein and Lazar, 2016).

## RESULTS

### Baseline Characteristics of Students Within and Between Conditions

There were no differences between the two groups in terms of student's age [ $t_{(38)} = 0.56$ ,  $p = 0.58$ ] or gender [ $X^2_{(1)} = 0.10$ ,  $p = 0.50$ ]. Table 2 shows an overview of these demographic characteristics as well as of some other potentially relevant characteristics. Results of *t*-tests on potential differences between conditions indicate that students in the two groups were comparable with respect to their writing knowledge [ $t_{(38)} = 0.77$ ,  $p = 0.44$ ], self-efficacy for ideation [ $t_{(38)} = 0.75$ ,  $p = 0.46$ ], self-efficacy for conventions [ $t_{(38)} = 1.01$ ,  $p = 0.32$ ], self-efficacy for the regulation of writing [ $t_{(38)} = -0.34$ ,  $p = 0.74$ ], and self-efficacy for five-paragraph essay writing [ $t_{(38)} = -0.46$ ,  $p = 0.65$ ].

### Quality of Peer Assessment

Results for the peer assessment phase show that the reliability and validity of the peer evaluations in the two conditions were quite comparable. Students evaluated the texts reliably within conditions, with an SSR of 0.83 for the pairwise comparisons and an average intraclass correlation coefficient 0.80 for the criteria judgments (ranging from 0.56 and 0.61 for the subdimensions grammar and vocabulary to 0.75 and 0.84 for respectively content/structure and coherence). In addition, evaluations between conditions correlated highly,  $r = 0.87$ ,  $p < 0.01$ .

There were considerable differences between conditions in how many evaluations students made during the 30-min time frame. Students in the pairwise comparisons condition made faster decisions (2.9 min for a comparison) than students in the criteria condition (5.6 min for a single essay). As a result, students in the CJ condition generally evaluated more essays than the students in the criteria condition. On average, students in the criteria condition evaluated only five of the ten essays (min = 2, max = 9), whereas students in the CJ condition completed ten comparisons (min = 3, max = 20). As one comparison includes

two essays, students in the CJ condition evaluated each essay twice on average.

As a result of evaluating more essays, CJ students provided also more than twice as much feedback than students in the criteria condition: 203 vs. 106 comments. There was no effect of condition on the number of aspects students commented on per essay, neither for positive feedback ( $p = 0.36$ ), nor for negative feedback ( $p = 0.41$ ). In both conditions, half of the feedback comments were focused on only one aspect of the essay at a time. For positive feedback at least two aspects were mentioned in 30 percent of the cases, with a maximum of 4 aspects per comment. For negative feedback this percentage was somewhat lower: only 23 percent, with a maximum of 3 aspects per comment. In the other cases the feedback segment was left blank by the student.

An in-depth analysis of the content of feedback showed considerable differences in the probability that a particular aspect was mentioned, see Tables 3 and 4 for an overview of the results. Table 3 shows that there were no significant differences between conditions for the proportion of positive feedback. The results were rather different for negative feedback, in which the condition affected the proportion of feedback in three of the five categories, see Table 4. Figure 2 shows the results for positive (left pane) and negative feedback (right pane) in more comprehensible terms: the proportion of feedback for each of the five categories. Below, these results of positive and negative feedback per feedback category are systematically presented.

First, when providing positive feedback, students in both conditions were equally likely to provide feedback on the content

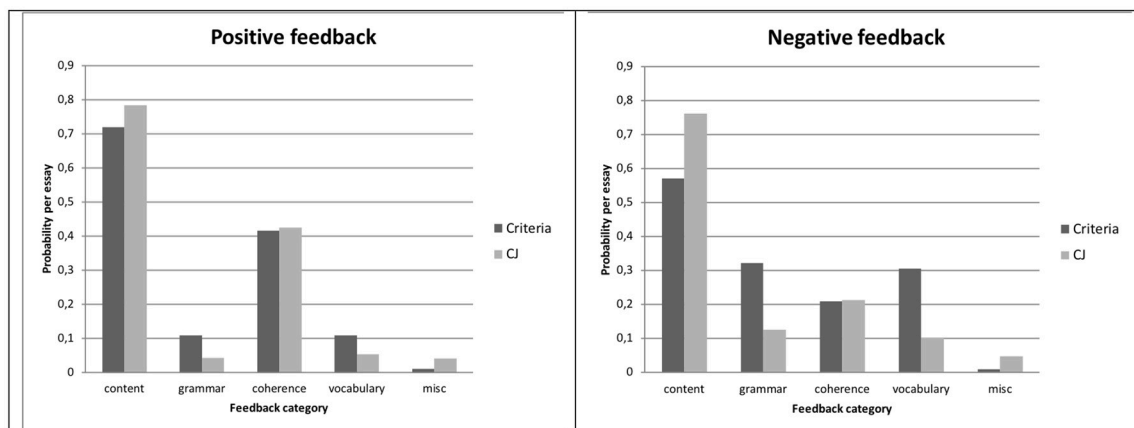
TABLE 2 | Characteristics of Students by Condition.

|                                 | Means (SD)                             |                                  |
|---------------------------------|--|----------------------------------|
|                                 | Criteria condition<br>( <i>n</i> = 20) | CJ condition<br>( <i>n</i> = 20) |
| Age                             | 19.45 (0.89)                           | 19.30 (0.80)                     |
| Writing knowledge               | 5.55 (2.46)                            | 4.90 (2.85)                      |
| Self-efficacy for ideation      | 63.75 (9.40)                           | 61.70 (7.89)                     |
| Self-efficacy for conventions   | 70.95 (10.95)                          | 67.65 (9.72)                     |
| Self-efficacy for regulation    | 60.33 (12.32)                          | 61.54 (10.17)                    |
| Self-efficacy for essay writing | 61.50 (14.52)                          | 63.50 (12.78)                    |

TABLE 3 | Estimates of Logistic Cross-Classified Multilevel Models for Positive Feedback by Category.

| Category              | Logit (SE)      |  |                        |                      |
|-----------------------|-----------------|--|------------------------|----------------------|
|                       | Intercept       | Condition<br>(0 = criteria;<br>1 = CJ) | S <sup>2</sup> student | S <sup>2</sup> essay |
| Content and structure | 0.95 (0.27)**   | 0.35 (0.35)                            | 0.23 (0.23)            | 0.00                 |
| Grammatical control   | -2.11 (0.46)*** | -0.99 (0.57)                           | 0.75 (0.62)            | 0.47 (0.44)          |
| Coherence and unity   | -0.34 (0.36)    | 0.04 (0.42)                            | 0.89 (0.40)**          | 0.22 (0.20)          |
| Vocabulary            | -2.10 (0.47)*** | -0.78 (0.60)                           | 1.13 (0.72)            | 0.27 (0.40)          |
| Miscellaneous         | -4.47 (0.97)*** | 1.31 (0.77)                            | 0.91 (0.80)            | 3.49 (2.32)          |

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .



**FIGURE 2 |** Effects of condition (criteria vs. CJ) on the estimated probability of positive feedback (left pane) and negative feedback (right pane) for each of the five categories: content and structure, grammatical control, coherence and unity, vocabulary, and miscellaneous.

**TABLE 4 |** Estimates of Logistic Cross-Classified Multilevel Models for Negative Feedback by Category.

| Category              | Logit (SE)      |   |                        |                      |
|-----------------------|-----------------|---|------------------------|----------------------|
|                       | Intercept       | Condition<br>(0 = criteria;<br>1 = C-J) | S <sup>2</sup> student | S <sup>2</sup> essay |
| Content and structure | 0.29 (0.31)     | 0.88 (0.33)**                           | 0.19 (0.22)            | 0.30 (0.25)          |
| Grammatical control   | −0.75 (0.30)*   | −1.19 (0.40)**                          | 0.41 (0.32)            | 0.09 (0.18)          |
| Coherence and unity   | −1.32 (0.39)**  | 0.02 (0.38)                             | 0.29 (0.32)            | 0.59 (0.45)          |
| Vocabulary            | −0.82 (0.31)**  | −1.35 (0.36)***                         | 0.00                   | 0.36 (0.31)          |
| Miscellaneous         | −4.56 (1.11)*** | 1.59 (1.21)                             | 2.14 (1.15)*           | 0.43 (0.63)          |

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

and structure of the text, with a probability of 0.72 ( $t = -3.51$ ,  $p < 0.01$ ). There was no significant effect of condition ( $t = 1.01$ ,  $p = 0.32$ ), and there were no significant differences between students (Wald  $z = 1.02$ ,  $p = 0.15$ ) and essays (redundant). In contrast, when students commented on weaknesses in the text, there was a large effect of condition on the probability of content and structure feedback ( $t = 2.65$ ,  $p < 0.01$ ). Students in the criteria condition commented on these kinds of aspects only half of the time (proportion = 0.57), whereas students in the CJ condition focused on these aspects in 76% of the cases. There were no significant differences between students (Wald  $z = 0.88$ ,  $p = 0.19$ ) and essays (Wald  $z = 1.19$ ,  $p = 0.12$ ).

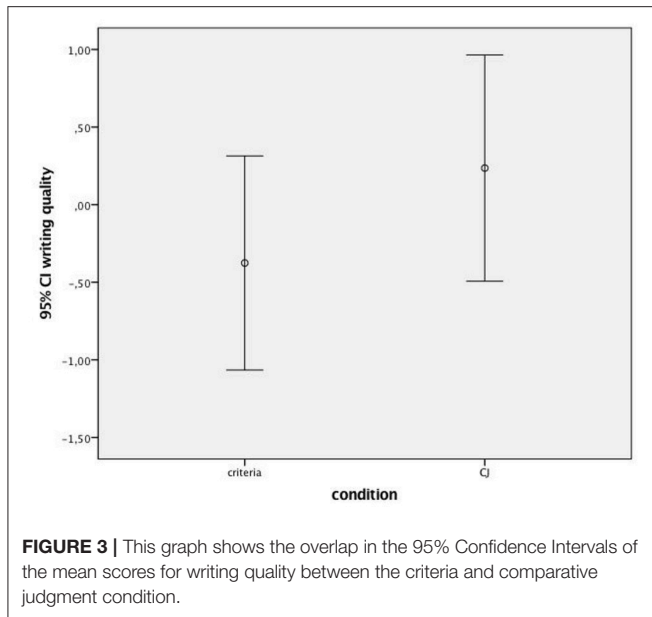
Second, the proportion of positive feedback on aspects related to grammatical control in the criteria condition was 0.11 ( $t = -4.58$ ,  $p < 0.001$ ). Although the proportion of feedback on grammar decreased in the CJ condition to only 0.04, this difference was only marginally significant ( $t = -1.73$ ,  $p = 0.09$ ). There were no significant differences between students (Wald  $z = 1.20$ ,  $p = 0.11$ ) and essays (Wald  $z = 1.08$ ,  $p = 0.14$ ). When students provided negative feedback, the proportion of feedback on grammar was not only higher, but there was also a negative

effect of condition ( $t = -1.19$ ,  $p < 0.01$ ). The proportion of grammar feedback in the criteria condition was 0.32, whereas in the CJ condition this was only 0.13. There were no significant differences between students (Wald  $z = 1.30$ ,  $p = 0.10$ ) and essays (Wald  $z = 0.49$ ,  $p = 0.32$ ).

Third, the probability of feedback on coherence and unity (0.42) was not significantly different from 0.50 ( $t = -0.95$ ,  $p = 0.34$ ), indicating that when students described strengths in a text, they commented on aspects that were related to coherence and unity half of the time. There were, however, large differences between students (Wald  $z = 2.25$ ,  $p < 0.05$ ): some students hardly provided feedback on coherence, whereas other students focused on coherence in more than three quarters of the cases {80% CI [0.13, 0.77]}. There was no effect of condition ( $t = 0.09$ ,  $p = 0.93$ ) and there were no significant differences between essays (Wald  $z = 1.09$ ,  $p = 0.28$ ). When students commented on weaknesses in the text, the probability that they focused on coherence and unity in the text was only 0.21. There was no effect of condition ( $t = 0.05$ ,  $p = 0.96$ ), and there were no significant differences between students (Wald  $z = 0.91$ ,  $p = 0.18$ ) and essays (Wald  $z = 1.32$ ,  $p = 0.08$ ).

Fourth, the results for feedback on vocabulary are quite comparable to the results for feedback on grammar. The proportion of positive feedback on vocabulary was 0.11 ( $t = -4.51$ ,  $p < 0.001$ ). There was no effect of condition ( $t = -1.29$ ,  $p = 0.20$ ), and there were no significant differences between students (Wald  $z = 1.57$ ,  $p = 0.06$ ) and essays (Wald  $z = 0.69$ ,  $p = 0.49$ ). In contrast, the proportion of negative feedback on grammar was generally higher, with a negative effect of condition as well ( $t = -3.79$ ,  $p < 0.001$ ). Specifically, the proportion of grammar feedback in the criteria condition was 0.31, compared to 0.10 for the students in the CJ condition. There were no significant differences between students (redundant) and essays (Wald  $z = 1.16$ ,  $p = 0.13$ ).

Fifth, students in both conditions hardly provided feedback on aspects that could not be categorized in any of the other four evaluation criteria, with a probability of 0.01 in both categories. Although students in the CJ condition mentioned somewhat



more miscellaneous aspects, there were no significant differences between condition ( $t < 1.70$ ,  $p > 0.09$ ). Only for negative feedback there were significant differences between students (Wald  $z = 1.86$ ,  $p < 0.03$ ).

### Quality of Writing Performance

Results indicated that students in the CJ-based peer assessment condition wrote texts of higher quality ( $M = 0.24$ ,  $SD = 1.56$ ) than students in the criteria-based peer assessment condition ( $M = -0.38$ ,  $SD = 1.47$ ), see also **Figure 3**. The average scores are presented in logits, which represent the probability that a particular text is judged as being of higher quality than a random text from the same pool of texts. In other words, the probability on high-quality texts was generally higher for students in the CJ condition (0.56) than for students in the criteria condition (0.41). An independent  $t$ -test revealed that the effect in this sample was moderate (Cohens'  $d = 0.40$ ), but statistically non-significant,  $t_{(38)} = -1.28$ ,  $p = 0.21$ . An additional analysis of covariance in which the effect of condition on writing quality was controlled for students' knowledge and self-efficacy for writing provided equal results,  $F_{(1,33)} = 3.48$ ,  $p = 0.22$ ,  $R^2 = 0.20$ .

## DISCUSSION

The present study aimed to investigate the differential learning effects of an instructional approach in which students apply analytic teacher-designed criteria to the evaluation of essays written by peers vs. an instructional approach in which students evaluate by comparison. This was tested in a small-scale authentic classroom situation, showing some interesting and promising findings. First, there were no difference in the reliability and validity of the judgments students made in each of the two conditions, indicating that both types of peer assessments equally support students in making evaluative judgments of the quality of their peers' essays. However, there were some differences between

conditions in the content of the peer feedback they provided. Compared to the criteria condition, students in the comparative judgment condition focused relatively more on aspects that were related to the content and structure of the text, and less so on aspects that were related to grammar and vocabulary. This was only the case for feedback targeted to aspects that needed improvement. For feedback on strengths, there appeared to be no difference between conditions. A second important finding of this study is that there appeared to be only a moderate effect of condition on the quality of students' own writing. Students in the comparative judgment condition wrote texts of somewhat higher quality than the students in the criteria condition. This difference was not significant in this sample, but that can be due to the relatively small sample size (cf. Wasserstein and Lazar, 2016). A posterior power analysis indicates that at least 98 students are needed per condition to have 80% power for detecting the moderate sized effect of 0.40 when employing the criterion level of 0.05 for statistical significance.

Two main conclusions can be drawn from these results. First and foremost, the instructional approaches influence the aspects of the text to which students pay attention when providing feedback. Although students in this study were all primarily focused on the content and structure of the text, especially when they provided positive feedback, they were more directed toward the lower level aspects of the text when they needed to provide suggestions for improvement based on an analytic list of criteria. However, when comparing essays, students stayed focused on the higher order aspects when identifying aspects that needed improvement. This finding might be due to the holistic approach in the process of comparative judgment, which allow students to make higher level judgments regarding the essay's communicative effectiveness.

Although it is not necessarily a bad thing to provide feedback on lower level aspects, feedback on higher level aspects is generally associated with improved writing performance (Underwood and Tregidgo, 2006). By doing so, the feedback in the comparative judgment condition can be more meaningful for the feedback receiver. Ultimately, this can also have an effect on feedback givers themselves as the way they evaluate texts and diagnose strengths and weaknesses in a peer's work may have an important influence on how they conceptualize and regulate quality in their own writing (Nicol and Macfarlane Dick, 2006; Nicol et al., 2014).

Second, conclusions regarding the effect of instructional approach on student's own writing performance are somewhat harder to draw based on the results of the present study. Although students in the comparative judgment condition on average wrote texts of higher quality than students in the criteria condition, this was definitely not the case for all students. Even when controlled for individual writing knowledge and writing self-efficacy, differences in writing quality were still larger within conditions than between conditions. Moreover, as the present study took place in an authentic classroom situation constraining the number of participating students, and as it is not ethical to exclude students from possible learning opportunities, it was deliberately decided not to implement a control condition in which students completed the same writing task without being presented with examples. As a result, students in both



conditions actively engaged with a range of examples of varying quality. As this process seems to be a necessary condition for students to develop a mental representation of what constitutes quality (Lin-Siegler et al., 2015; Tai et al., 2017), it could very well be the case that students in both conditions significantly improved their writing. More research is needed to examine whether the active use of shared criteria and examples in a peer assessment affects students' learning and performance, above and beyond the instructional approach (teacher-designed criteria or comparative judgment). Another opportunity for further research is to investigate how many examples of which quality are necessary for students to learn.

A possible explanation for the small effects in this study of the learning by comparison condition on students' writing quality may be that improved understanding of writing quality does not easily transfer to one's own writing, at least not on the short time. Further research is needed to understand what instructional factors can foster this transfer. For instance, the learning effects might be stronger once the peer assessment is routinely and systematically implemented in the curriculum. According to Sadler (1998), any feedback-enhanced intervention in which students are engaged in the process of assessing quality must be carried out long enough for it will be viewed by learners as normal and natural (p. 78). To our knowledge, there is no research yet that investigates how the number of peer assessments performed over the course of a curriculum affects students' performance.

The role of the teacher in the transfer from understanding to performance may be a crucial factor as well. Key aspects of pedagogical interventions that successfully promote student's learning include a combination of direct instruction, modeling, scaffolding and guided practice (Merrill, 2002). This implies that a peer assessment on its own may not be sufficient to improve writing. A more effective implementation of any type of peer assessment may be that teachers discuss the results from the peer assessment with students and show how they can use the information from the peer assessment during their own writing process (Sadler, 1998, 2009; Rust et al., 2003; Hendry et al., 2011, 2012; Carless and Chan, 2017). This may be especially true for comparative judgment in which students gradually develop their own understanding of criteria and standards for writing quality through comparing a range of texts from low to high quality, but without any explicit information and/or teacher guidance on the accuracy of their internally constructed standard of quality. At the end of the present study, students in comparative judgment condition confirmed that they missed explicit clues on whether they made the right choices during their comparisons. While acknowledging the importance of teachers, Sadler (2009) remarks that teachers should hold back from being too directive in guiding students' learning process. He states that students assume that teachers are the only agents who can provide effective feedback on their work and that they need a considerable period of practice and adaptation to build trust in the feedback they give and receive from peers, especially when they do this in a more holistic manner. When teachers are too directive in this procedure and keep focusing on analytic criteria instead of on the quality of texts as a whole, students' own learning process might be inhibited. Instead he argues that teachers should guide the process more indirectly, for instance, through monitoring students' evaluation

process from a distance and by providing meta-feedback on the quality of students' peer feedback. Together, this implies that a combination of both instructional methods might be more effective than either of them, and that teachers play an important role in how to bring criteria and examples together in such a way that students engage in deep learning processes.

Although the present study provides important insights into how students evaluate work of their peers and what aspects they take into account during these evaluations, the results do not provide any insight into how they evaluate their own work during writing. Theories on evaluative judgment suggest that improved understanding of what constitutes quality does not only improve how students evaluate the work of their peers but also how they evaluate their own work (Boud, 2000; Tai et al., 2017). Although writing researchers have already acquired a decent understanding of how novice and more advanced writers plan their writing product, there is not much information yet on how students evaluate and revise their writing. This is especially relevant for developing writers, as being able to monitor and control the quality of one's own product during writing is one of the most important predictors of writing quality (Flower and Hayes, 1980). Based on the small effects of peer assessment on writing quality in this research it might very well be possible that students have made changes in their writing process. To further our understanding of the learning effects of peer assessment in the context of writing, research should therefore take into account both the process and the product of writing.

## CONCLUSION

To summarize, the present study has taken a first but promising step into unraveling how analyzing examples of varying quality might foster students' understanding and performance in writing. It has been demonstrated that students analyze example texts quite differently by comparison than by applying teacher-designed criteria. In particular, when providing feedback in a comparative approach, students focus more on higher level aspects in their peers' texts. Although the results are not conclusive in whether the effects of learning by comparison also transfer to students' own writing performance, the results do suggest that it can be a powerful instructional tool in today's practice. It inherently activates students to engage with a range of examples of varying quality, doing so in a highly feasible and efficient manner (cf. Bartholomew et al., 2018a). Follow-up research is needed to really get a grip on the potential learning effects of comparative judgment, both to contrast the effects to other instructional approaches such as linking example texts to analytic criteria which is now regularly used in educational practice, but also with regards to contextual factors that are needed for an optimal implementation in practice.

## ETHICS STATEMENT

This study was carried out in accordance with the guidelines of the University of Antwerp. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## REFERENCES

- Bartholomew, S. R., Nadelson, L. S., Goodridge, W. H., and Reeve, E. M. (2018a). Adaptive comparative judgment as a tool for assessing open-ended design problems and model eliciting activities. *Edu. Assess.* 23, 85–101. doi: 10.1080/10627197.2018.1444986
- Bartholomew, S. R., Strimel, G. J., and Yoshikawa, E. (2018b). Using adaptive comparative judgment for student formative feedback and learning during a middle school design project. *Int. J. Technol. Des. Edu. Adv.* 1–3. doi: 10.1007/s10798-018-9442-7
- Bell, A., Mladenovic, R., and Price, M. (2013). Students' perceptions of the usefulness of marking guides, grade descriptors and annotated exemplars. *Assess. Eval. Higher Edu.* 38, 769–788. doi: 10.1080/02602938.2012.714738
- Bloxham, S., and Campbell, L. (2010). Generating dialogue in assessment feedback: exploring the use of interactive cover sheets. *Assess. Eval. Higher Edu.* 35, 291–300. doi: 10.1080/02602931003650045
- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies Contin. Edu.* 22, 151–167. doi: 10.1080/713695728
- Bouwer, R., Béguin, A., Sanders, T., and Van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Lang. Test.* 32, 83–100. doi: 10.1177/0265532214542994
- Bouwer, R., Goossens, M., Mortier, A. V., Lesterhuis, M., and De Maeyer, S. (2018). “Een comparatieve aanpak voor peer assessment: Lereren door te vergelijken [A comparative approach for peer assessment: Learning by comparison],” in *Toetsrevolutie: Naar een feedbackcultuur in het hoger onderwijs [Assessment Revolution: Towards a Feedback Culture in Higher Education]*, eds D. Sluijsmans and M. Segers (Culemborg: Phonese), 92–106.
- Bradley, R. A., and Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* 39, 324–345. doi: 10.1093/biomet/39.3.4324
- Brookhart, S. M. (2018). Appropriate criteria: key to effective rubrics. *Front. Edu.* 3:22. doi: 10.3389/educ.2018.00022
- Bruning, R., Dempsey, M., Kauffman, D. F., McKim, C., and Zumbrunn, S. (2013). Examining dimensions of self-efficacy for writing. *J. Edu. Psychol.* 105, 25–38. doi: 10.1037/a0029692
- Carless, D., and Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assess. Eval. Higher Edu.* doi: 10.1080/02602938.2018.1463354
- Carless, D., and Chan, K. K. H. (2017). Managing dialogic use of exemplars. *Assess. Eval. Higher Edu.* 42, 930–941. doi: 10.1080/02602938.2016.1211246
- D-PAC [Computer software] (2018). *D-PAC [Computer software]*. Antwerp: University of Antwerp. Available online at: <http://www.d-pac.be>.
- Fielding, A., and Goldstein, H. (2006). *Cross-Classified and Multiple Membership Structures in Multilevel Models: an Introduction and Review (Report no. 791)*. London: DFES.
- Flower, L., and Hayes, J. (1980). “The dynamics of composing: making plans and juggling constraints,” in *Cognitive Processes in Writing*, eds L. Gregg and E. Steinberg (Hillsdale, NJ: Erlbaum), 3–30.
- Fraile, J., Panadero, E., and Pardo, R. (2017). Co-creating rubrics: the effects on self-regulated learning, self-efficacy and performance of establishing assessment criteria with students. *Stud. Edu. Eval.* 53, 69–76. doi: 10.1016/j.stueduc.2017.03.003
- Gentner, D., Loewenstein, J., and Thompson, L. (2003). Leaning and transfer: a general role for analogical encoding. *J. Edu. Psychol.* 95, 393–408. doi: 10.1037/0022-0663.95.2.393
- Gill, T., and Bramley, T. (2013). How accurate are examiners' holistic judgments of script quality? *Assess. Edu.* 20, 308–324. doi: 10.1080/0969594X.2013.779229
- Handley, K., and Williams, L. (2011). From copying to learning: using exemplars to engage students with assessment criteria and feedback. *Assess. Eval. Higher Edu.* 36, 95–108. doi: 10.1080/02602930903201669
- Hendry, G. D., Armstrong, S., and Bromberger, N. (2012). Implementing standards-based assessment effectively: incorporating discussion of exemplars into classroom teaching. *Assess. Eval. Higher Edu.* 37, 149–161. doi: 10.1080/02602938.2010.515014
- Hendry, G. D., Bromberger, N., and Armstrong, S. (2011). Constructive guidance and feedback for learning: the usefulness of exemplars, marking sheets and different types of feedback in a first year law subject. *Assess. Eval. Higher Edu.* 36, 1–11. doi: 10.1080/02602930903128904
- Huisman, B., Saab, N., van Driel, J., and van den Broek, P. (2018). Peer feedback on academic writing: undergraduate students' peer feedback role, peer feedback perceptions and essay performance. *Assess. Eval. Higher Edu.* 43, 955–968. doi: 10.1080/02602938.2018.1424318
- Jones, I., and Alcock, L. (2014). Peer assessment without assessment criteria. *Stud. Higher Edu.* 39, 1774–1787. doi: 10.1080/03075079.2013.821974
- Jones, I., and Wheadon, C. (2015). Peer assessment using comparative and absolute judgment. *Stud. Edu. Eval.* 47, 93–101. doi: 10.1016/j.stueduc.2015.09.004
- Kok, E. M., de Bruin, A. B. H., Robben, S. G. F., and van Merriënboer, J. J. G. (2013). Learning radiological appearances of diseases: does comparison help? *Learn. Instruct.* 23, 90–97. doi: 10.1016/j.learninstruc.2012.07.004
- Laming, D. R. J. (2004). *Human Judgment: The Eye of the Beholder*. London: Thomson Learning.
- Lin-Siegler, X., Shaenfield, D., and Elder, A. D. (2015). Contrasting case instruction can improve self-assessment of writing. *Educ. Technol. Res. Dev.* 63, 517–537. doi: 10.1007/s11423-015-9390-9
- Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York, NY: John Wiley.
- Merrill, M. D. (2002). First principles of instruction. *Edu. Technol. Res. Dev.* 50, 43–59. doi: 10.1007/BF02505024
- Nicol, D., Thomson, A., and Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assess. Eval. Higher Edu.* 39, 102–122. doi: 10.1080/02602938.2013.795518
- Nicol, D. J., and Macfarlane Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Stud. Higher Edu.* 31, 199–218. doi: 10.1080/03075070600572090
- Nuzzo, R. (2014). Scientific method: statistical errors. *Nature* 506, 150–152. doi: 10.1038/506150a
- Orsmond, P., Merry, S., and Reiling, K. (2002). The use of exemplars and formative feedback when using student derived marking criteria in peer and self-assessment. *Assess. Eval. Higher Edu.* 27, 309–323. doi: 10.1080/0260293022000001337
- Patchan, M., Charney, D., and Schunn, C. D. (2009). A validation study of students' end comments: comparing comments by students, a writing instructor, and a content instructor. *J. Writing Res.* 1, 124–152. doi: 10.17239/jowr-2009.01.02.2
- Patton, C. (2012). Some kind of weird, evil experiment: student perceptions of peer assessment. *Assess. Eval. Higher Edu.* 37, 719–731. doi: 10.1080/02602938.2011.563281
- Peng, C. Y. J., Lee, K. L., and Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *J. Edu. Res.* 96, 3–14. doi: 10.1080/00220670209598786
- Pollitt, A. (2004). “Let's stop marking exams,” *Paper Presented at the Conference International Association of Educational Assessment* (Philadelphia, PA).
- Rust, C., Price, M., and O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assess. Eval. Higher Edu.* 28, 147–164. doi: 10.1080/0260293032000045509

## FUNDING

This work was supported by the Flanders Innovation & Entrepreneurship and the Research Foundation [Grant No. 130043].

- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instruct. Sci.* 18, 119–144.
- Sadler, D. R. (1998). Formative assessment: revisiting the territory. *Assess. Edu. Princ. Pol. Prac.* 5, 77–84.
- Sadler, D. R. (2002). “Ah!... So That’s “Quality,” in *Assessment: Case Studies, Experience and Practice from Higher Education*, eds P. Schwartz and G. Webb (London: Kogan Page), 130–136.
- Sadler, D. R. (2009). “Transforming holistic assessment and grading into a vehicle for complex learning,” in *Assessment, Learning and Judgment in Higher Education*, ed G. Joughin (Dordrecht: Springer), 1–19.
- Sadler, D. R. (2010). Beyond feedback: developing student capability in complex appraisal. *Assess. Eval. Higher Edu.* 35, 535–550. doi: 10.1080/02602930903541015
- Tai, J., Ajajawi, R., Boud, D., Dawson, P., and Panadero, E. (2017). Developing evaluative judgment: enabling students to make decisions about the quality of work. *Higher Edu.* 13, 1–15. doi: 10.1007/s10734-017-0220-3
- To, J., and Carless, D. (2016). Making productive use of exemplars: peer discussion and teacher guidance for positive transfer of strategies. *J. Further Higher Educ.* 40, 746–764. doi: 10.1080/0309877X.2015.1014317
- 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.* 14, 281–294. doi: 10.1080/09695940701591867
- Underwood, J. S., and Tregidgo, A. P. (2006). Improving student writing through effective feedback: best practices and recommendations. *J. Teach. Writing* 22, 73–98.
- Verhavert, S., De Maeyer, S., Donche, V., and Coertjens, L. (2017). Scale separation reliability: what does it mean in the context of comparative judgment? *Appl. Psychol. Meas.* doi: 10.1177/0146621617748321
- Wasserstein, R. L., and Lazar, N. A. (2016). The ASA’s statement on p-values: context, process, and purpose. *Am. Statist.* 70, 129–133. doi: 10.1080/00031305.2016.1154108.
- Yucel, R., Bird, F. L., Young, J., and Blanksby, T. (2014). The road to self-assessment: exemplar marking before peer review develops first-year students’ capacity to judge the quality of a scientific report. *Assess. Eval. Higher Edu.* 39, 971–986. doi: 10.1080/02602938.2014.880400
- Zimmerman, B. J., and Risemberg, R. (1997). Becoming a self-regulated writer: a social cognitive perspective. *Contempor. Educat. Psychol.* 22, 73–101.

**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 Bouwer, Lesterhuis, Bonne and De Maeyer. 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.

## APPENDIX

**TABLE A1 |** International Trade English 2A – writing class  
**Evaluation grid 5/paragraph essay:**

|   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
| <b>Structure and content:</b>   |   |   |   |   |   |   |   |
| - relevant title  | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| - introduction: motivator, topic sentence, road map   |   |   |   |   |   |   |   |
| - body : subtopic sentences   |   |   |   |   |   |   |   |
| - conclusion : no new info, reworded topic sentence, link to motivator                        |   |   |   |   |   |   |   |
| - visual structure  |   |   |   |   |   |   |   |
| - logical structure   |   |   |   |   |   |   |   |
| - selected content is relevant/valuable for business students                                 |   |   |   |   |   |   |   |
| <b>Grammatical control:</b>   |   |   |   |   |   |   |   |
| - no horrors of English   |   |   |   |   |   |   |   |
| o 1HoE = -2   |   |   |   |   |   |   |   |
| o 2HoE = -3   | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| o 3HoE = -4   |   |   |   |   |   |   |   |
| - low/high degree of lexicogrammatical accuracy (fluent sentences and correct use of grammar) |   |   |   |   |   |   |   |
| - no sloppy errors/typos/inconsistencies  |   |   |   |   |   |   |   |
| <b>Coherence:</b>   |   |   |   |   |   |   |   |
| - use of linking words  |   |   |   |   |   |   |   |
| - use of paraphrases (in subtopic sentences)  |   |   |   |   |   |   |   |
| - support is well-illustrated   |   |   |   |   |   |   |   |
| <b>Unity:</b>   |   |   |   |   |   |   |   |
| - one topic per paragraph   | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| - stays on topic during essay   |   |   |   |   |   |   |   |
| <b>Formal essay</b> (no I, we...)   |   |   |   |   |   |   |   |
| <b>Vocabulary:</b>  |   |   |   |   |   |   |   |
| - good range of vocabulary related to topic   | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| - specific and to the point   |   |   |   |   |   |   |   |
| - varied formulation  |   |   |   |   |   |   |   |

**Describe here the strengths and weaknesses of the essay. Be as specific as possible.**

**Strengths:**

**Weaknesses:**



# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read  
for greatest visibility  
and readership



## FAST PUBLICATION

Around 90 days  
from submission  
to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,  
and constructive  
peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers  
acknowledged by name  
on published articles

## Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

Visit us: [www.frontiersin.org](http://www.frontiersin.org)

Contact us: [info@frontiersin.org](mailto:info@frontiersin.org) | +41 21 510 17 00



## REPRODUCIBILITY OF RESEARCH

Support open data  
and methods to enhance  
research reproducibility



## DIGITAL PUBLISHING

Articles designed  
for optimal readership  
across devices



## FOLLOW US

@frontiersin



## IMPACT METRICS

Advanced article metrics  
track visibility across  
digital media



## EXTENSIVE PROMOTION

Marketing  
and promotion  
of impactful research



## LOOP RESEARCH NETWORK

Our network  
increases your  
article's readership