Check for
updates

Appropriate Criteria: Key to Effective Rubrics

Susan M. Brookhart*

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

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.

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). The 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 3 levels – 3 4 levels – 14 5 levels – 8 6 levels – 2 8 levels – 1 Total – 29	4 levels – 4 5 levels – 1 7 levels – 1 Total – 6	35
	Holistic	4 levels – 3 5 levels – 1 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 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 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 Research assistants ICC 0.77 Peers (4–5 per group) ICC 0.79 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 Detailed task-specific point schemes for each task	103 students in 3 courses (final version), 432 students in 14 courses during test development	$\kappa = 0.41$ consistency between criteria $\alpha = 0.82$
Cho et al., 2006	Undergraduate, graduate	SWoRD Writing Rubrics 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 Single rater ICCs 0.17–0.56 Multiple rater ICCs 0.45–0.88 Compared reliability from student and instructor perspectives
Ciorba and Smith, 2009	Undergraduate	Music – Instrumental and vocal performance 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 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 Single rater ICCs .33 −0.50 (individual criteria),0.50 −0.60 (total score)
Garcia-Ros, 2011	Undergraduate	Oral presentation 14 criteria organized into 4 areas. 4 levels (0-3) with descriptive plds	64 educational psychology students	exact agreement = 66% adjacent agreement = 98% $\kappa = 0.36$ exact agreement $\kappa = 0.80$ adjacent agreement median $r = 0.89$
Kocakülah, 2010	Undergraduate	Newton's Laws of Motion problem solving 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 single rater ICCs, 0.14, 0.38 multiple rater ICCs, 0.93, 0.98 instructor's consistency between 2 forms, median $\alpha = 0.76$
Lewis et al., 2008	Undergraduate	Acute care treatment planning Analytic rubric with 4 criteria and descriptive plds at 4 levels	22 students, 5 clinical educators, 1 academic faculty	Expert raters Single rater ICC = 0.32
Menéndez-Varela and Gregori-Giralt, 2016	Undergraduate	Service learning projects 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 α increased from 0.67 (at stage 2 of study) to 0.93 (at stage 3 of study; α for oral presentation skills was 0.77
Newman et al., 2009	Graduate faculty	Peer assessment of faculty teaching Rating scale, 1–5 (excellent through does not demonstrate criterion), on 11 criteria	14 resource faculty	Expert raters Single rater ICC = 0.27 (total score)
Nicholson et al., 2009	Undergraduate	Nurse clinical performance in operating suite 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 Single rater ICCs .51 −0.61 Multiple rater ICC =0.98
Pagano et al., 2008	Undergraduate	Writing (College composition) 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 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 Single rater ICCs 0.90–0.95 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 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 5 criteria (for each section of the paper) based on Facione and Facione (1996), with descriptive plds at5 levels	2002, 30 papers; 2003, 30 papers	Median r = 0.90
Schreiber et al., 2012	Undergraduate	Public Speaking Competence Rubric 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 Multiple rater ICCs 0.91, 0.93
Stellmack et al., 2009	Undergraduate	Writing APA-style introductions Analytic rubric with 8 criteria with descriptive plds at 4 levels	40 papers, 3 researcher/graders	Interrater agreement exact = 0.37, adjacent = 0.90 Intrarater agreement exact = 0.78, adjacent = 0.98 κ = 0.33
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	Generalizability for relative decisions = 0.85
Wald et al., 2012	Graduate	Reflective writing 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 Inter-judge consistency, median α = 0.77
Wallace et al., 2011	Undergraduate	Astronomy – Cosmology Task-specific, holistic rubrics for each test item, with 5 levels	65 responses from 21 students, 9 items	Exact agreement, overall score = 83% κ = 0.76, weighted κ = 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 Analytic rubric with 3 criteria, 3 levels with mostly descriptive plds	230 speeches (112 individual, 118 group)	Based on student learning outcomes; 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 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 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; Student interviews[Student results differed by course (could differentiate types of instruction), criterion-related evidence (to physics grades)
Cho et al., 2006	Undergraduate, graduate	Writing 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 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); Scale intercorrelations (internal validity evidence)
Garcia-Ros, 2011	Undergraduate	Oral presentation 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; Faculty perceptions
Jonsson, 2014	Graduate	3 rubrics Survey construction rubric in epidemiology: analytic, general rubric, 2 criteria, 4 levels with plds for each House inspection rubric in real estate program: more like a checklist, w/ multiple criteria and a tally of facts and reasoning for each 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). Results were interpreted to mean that rubrics made assessment expectations explicit for students.
Kocakülah, 2010	Undergraduate	Physics – Newton's Laws of Motion problems 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 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 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 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; Faculty perceptions
Pagano et al., 2008	Undergraduate	Writing (College composition) 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 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 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; Student perceptions
Rezaei and Lovorn, 2010	Graduate	Writing 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 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 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 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 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 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 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 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 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 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 Analytic rubric with 6 criteria and descriptive plds at 5 levels	Results summarized, sample size not given
Knight, 2006	Undergraduate	Information Literacy (Annotated Bibliographies) 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 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 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 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 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 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.
Kocakülah, 2010	Undergraduate	Newton's Laws of Motion problem solving 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 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 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 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)

Study	Level	Rubric topic & description	Sample	Design	Findings
Reynolds-Keefer, 2010	Undergraduate	Writing 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 "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 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 that 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

Kocakülah, 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., Garci,á, 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