

A review of rubric use in higher education

Y. Malini Reddy^{a*} and Heidi Andrade^b

^a*ICFAI Business School, Hyderabad, India;* ^b*Educational and Counseling Psychology, University at Albany, NY, USA*

This paper critically reviews the empirical research on the use of rubrics at the post-secondary level, identifies gaps in the literature and proposes areas in need of research. Studies of rubrics in higher education have been undertaken in a wide range of disciplines and for multiple purposes, including increasing student achievement, improving instruction and evaluating programmes. While, student perceptions of rubrics are generally positive and some authors report positive responses to rubric use by instructors, others noted a tendency for instructors to resist using them. Two studies suggested that rubric use was associated with improved academic performance, while one did not. The potential of rubrics to identify the need for improvements in courses and programmes has been demonstrated. Studies of the validity of rubrics have shown that clarity and appropriateness of language is a central concern. Studies of rater reliability tend to show that rubrics can lead to a relatively common interpretation of student performance. Suggestions for future research include the use of more rigorous research methods, more attention to validity and reliability, a closer focus on learning and research on rubric use in diverse educational contexts.

Keywords: rubric; higher education; formative assessment; review of empirical research; reliability and validity; user perceptions; effects on performance; programme assessment

Introduction

Educators tend to define the word ‘rubric’ in slightly different ways. A commonly used definition is a document that articulates the expectations for an assignment by listing the criteria or what counts, and describing levels of quality from excellent to poor (Andrade 2000; Stiggins 2001; Arter and Chappuis 2007). For an example of an analytic rubric that meets this definition and has been used in an undergraduate course on educational psychology, see Table 1.

A rubric has three essential features: evaluation criteria (the leftmost column in Table 1), quality definitions (the second, third and fourth columns in Table 1) and a scoring strategy (Popham 1997). Evaluation criteria are the factors that an assessor considers when determining the quality of a student’s work. Also described as a set of indicators or a list of guidelines, the criteria reflect the processes and content judged to be important (Parke 2001). Quality definitions provide a detailed explanation of what a student must do to demonstrate a skill, proficiency or criterion in order to attain a particular level of achievement, for example poor, fair, good or excellent. The quality definitions address the need to distinguish between good and poor responses, both

*Corresponding author. Email: malinireddyicfai@yahoo.co.in

Table 1. Example rubric from undergraduate course on educational psychology. Rubric for the annotated bibliography.

	A	B	C	D	E
Research topic, questions and relevance (3 pts.)	The paper provides a clear, focussed description of your research topic and question(s), and discusses the relevance of your topic to education.	The paper tells the topic and lists the questions. The discussion of relevance is very broad but appropriate.	Information about the topic, research question(s) and relevance is too vague, unclear or incomplete.	Little appropriate information is given about the topic, questions or relevance.	0
References (4 pts.)	The paper gives complete bibliographic information in APA style for five relevant journal papers. No more than one paper is from a website.	The paper gives complete bibliographic information for five relevant journal papers but some or all citations are incorrectly formatted.	The paper gives incomplete or incorrect bibliographic information for five journal papers, some of which are not appropriate.	The paper gives bibliographic information for < five relevant journal papers.	0
Annotations (7 pts.)	In one page or less, the annotations identify the background or authority of the authors, summarise major themes and explain how the paper addresses your research questions.	The annotations summarise each paper but did not identify authority and/or make explicit connections to the research questions.	The annotations are either retellings of the paper, or unclear, or incomplete or off topic.	The annotations are token.	0
Implications for practice (7 pts.)	Each annotation notes practical implications for teaching that follow logically from the paper.	Practical implications are listed but more should have been noted. One or two may not follow logically from the papers.	The implications for teaching are limited, vague, distantly related to and/or in conflict with the literature.	Little accurate information is provided about teaching practices.	0
Conventions (4 pts.)	The paper has page numbers, is typed, double-spaced and well organised. All ideas are clearly articulated and carefully cited. No problems with paragraph format, spelling, punctuation, grammar, etc. Readable paper copies are attached.	A few problems with organisation, clarity or conventions should have been fixed but are not serious enough to distract the reader. Copies of each article are attached to the paper.	Numerous errors are distracting but do not interfere with meaning. Copies of some articles are attached.	Frequent problems make the paper hard to understand. Possible plagiarism risks the appearance of cheating. No copies.	0

for scoring purposes and to provide feedback to students. Scoring strategies for rubrics involve the use of a scale for interpreting judgments of a product or process. Scoring strategies will not be discussed here because the calculation of final grades is not a concern of this review.

Rubrics are often used by teachers to grade student work but many authors argue that they can serve another, more important, role as well: When used by students as part of a formative assessment of their works in progress, rubrics can teach as well as evaluate (Arter and McTighe 2001; Stiggins 2001). Used as part of a student-centered approach to assessment, rubrics have the potential to help students understand the targets for their learning and the standards of quality for a particular assignment, as well as make dependable judgments about their own work that can inform revision and improvement.

There is limited but compelling empirical evidence that rubric use can promote learning and achievement by primary and secondary students (Cohen et al. 2002; Andrade, Du, and Wang 2008). The purposes of this review are to examine the type and extent of empirical research on rubrics at the post-secondary level and to stimulate research on rubric use in post-secondary teaching. The questions that motivated this review include: (1) In general, what kind of research has been done on rubrics in higher education? (2) More specifically, is there evidence that rubrics can be used as formative assessments in order to promote learning and achievement in higher education, as opposed to rubric use that serves only the purposes of grading and accountability? (3) How much attention has been paid to the quality of the rubrics being used by college and university instructors? and (4) What are some fruitful directions for future research and development in this area?

The 20 articles included in the review were retrieved online using two inclusion criteria: 'empirical research' and 'higher education'. Master's theses about rubrics were excluded from the review, though they were numerous. Doctoral dissertations were included if they appeared to use research methods that could lead to credible results. The keywords searched include rubric, higher education, post-secondary and empirical studies. The databases searched include ABI Inform Global, Academic Search Premier, Blackwell Journals, CBCA Education, Sage Journals Online, Emerald Full Text, SpringerLink, ProQuest Education Journals, ERIC, J-Store, PsychInfo, Education Research Complete and EBSCO.

Overview of rubric use in higher education

The literature suggests that rubrics are being used in a wide variety of disciplines in higher education, if not by a large number of instructors. The disciplines for which there are published studies of rubric use include the liberal arts, information literacy, medicine, nursing, management, dentistry, food technology, teacher education and film technology. Rubrics are used to provide feedback on and to grade an array of student products, including concept maps, literature reviews, reflective writings, bibliographies, oral presentations, critical thinking, citation analyses, portfolios, projects and oral and written communication skills.

While some studies (e.g. Song 2006) delve into how the diagnostic feedback gained by the use of a rubric can be used to identify areas for improvement in instruction, other researchers use rubrics solely to evaluate student work (Campbell 2005; Tunon and Brydges 2006). Andrade (2005) has argued that rubrics can serve instructional as well as evaluative purposes. Several researchers report such uses and provide

evidence that rubrics support teaching and learning (Powell 2001; Osana and Seymour 2004; Reitmeier, Svendsen, and Vrchota 2004; Andrade and Du 2005; Schneider 2006). Rubrics are also being used in programme evaluation (Dunbar, Brooks, and Kubicka-Miller 2006; Knight 2006; Oakleaf 2006).

Research on the results of rubric use

The studies of rubric use at the post-secondary level included in this review have been organised according to the overarching themes that were detected by noting topics that recurred across studies. The themes include student and instructor perceptions of rubric use, the effect of rubrics on learning or academic performances, the use of rubrics as instructional and programme assessments and studies of validity and reliability.

Student perceptions of rubric use

Studies of students' responses to rubric use suggest that graduate and undergraduate students value rubrics because they clarify the targets for their work, allow them to regulate their progress and make grades or marks transparent and fair. The undergraduate and graduate business students ($N = 150$) in Bolton's (2006) study asserted that rubrics enabled them to engage in important processes, including identifying critical issues in an assignment and, thereby, reducing uncertainty and doing more meaningful work, determining the amount of effort needed for an assignment, evaluating their own performances in order to get immediate feedback, especially on weaknesses, estimating their grades prior to the submission of assignments and focussing their efforts so as to improve performance on subsequent assignments.

These findings are strikingly similar to those found by Andrade and Du (2005). The 14 pre-service teacher education undergraduates interviewed for this study reported that they used rubrics to plan an approach to an assignment, check their work and reflect on feedback from others. They said that using rubrics helped them focus their efforts, produce work of higher quality, earn better grades and feel less anxious about an assignment. The students also emphasised their perceptions of rubric-referenced grading as fair and transparent.

A third study, with similar findings, was conducted by Powell (2001). In a three cycle action research inquiry, rubrics were implemented for assessing creative media projects. Qualitative analysis based on questionnaires and interviews revealed that providing rubrics when handing out and explaining the assignment prompt or brief, as well as the use of rubrics for grading or marking by the instructor was associated with better student attitudes about fairness and satisfaction with grading.

The students in the three studies described above had been involved in courses in which rubrics were used as part of a formative, student-centered approach to assessment. The fact that rubrics were either co-created with or available to students before they began an assignment is probably the key to understanding the students' positive responses to them. This is suggested by a study by Schneider (2006), which involved introducing two rubrics to undergraduate education students ($N = 55$) at different points in the assessment process. While, the first was made available to students only after it had been used for grading an assignment, the second rubric was distributed along with the assignment brief. This single group action research used a Likert-type survey and open ended questions to collect students' initial reactions to rubric use as well as their experience of using the first and second rubrics. Schneider found that all

the students wanted to use rubrics again. However, the rubric that was handed out with the assignment was rated as useful by 88% of the students, as compared to the 10% who found the rubric useful when it was provided only with a final grade. For these and other reasons, researchers stress the instructional value of rubrics and urge instructors to use them as instructional guides, not just grading tools (Andrade 2000; Osana and Seymour 2004; Tierney and Simon 2004; Song 2006).

Instructor perceptions of rubric use

There is evidence of both positive responses and resistance to rubric use by college and university instructors. Three studies report positive instructor perceptions of rubrics as scoring guides. The instructor in Powell's (2001) study of assessment in film and television production courses felt that rubrics provided an objective basis for evaluation. Likewise, Campbell (2005), who described the development of a rubric-based e-marking tool, reported that the instructors who used the e-marking tool with multiple classes perceived that the tool made them mark or grade more consistently, reliably and efficiently. A similar finding is from Reitmeier, Svendsen, and Vrchota, who reported that the use of rubrics for oral presentation evaluations facilitated the change in evaluation procedures from 'subjective observations to specific performances' (2004, 18).

Parkes (2006), however, reports contrasting results. Employing a two-group pre-post-quasi-experimental research design, Parkes examined the impact of music performance rubrics on grading satisfaction by students ($N = 44$) and faculty ($N = 11$). Conducted in three music institutions, the study did not find any significant differences in student and teacher attitudes towards grading after the use of rubrics. Noting that the unwillingness of the faculty to participate was a major limitation of the study, Parkes emphasised the need to better understand the reasons for faculty hesitation regarding the study and use of rubrics.

One striking difference between students' and instructors' perceptions of rubric use is related to their perceptions of the purposes of rubrics. Students frequently referred to them as serving the purposes of learning and achievement, while instructors focussed almost exclusively on the role of a rubric in quickly, objectively and accurately assigning grades. Instructors' limited conception of the purpose of a rubric might contribute to their unwillingness to use them. College and university teachers might be more receptive if they understand that rubrics can be used to enhance teaching and learning as well as to evaluate.

Rubric use and academic performance

The linkage between rubrics and learning has been explored by several researchers, with results generally suggesting higher achievement and deeper learning by students who have rubrics to guide their work. Petkov and Petkova (2006) examined the final grades for a short-term project in a post-graduate level course entitled Management of Business Information. The two-group post-test design involved assigning a similar project to two classrooms with 20 students each. The students in one of the classrooms were provided with the project rubric at the beginning of the semester. Comparison of the project grades attained by students in the two classrooms showed that the mean percentage grade for the section using rubrics was significantly higher than the comparison group of students.

Similarly, Reitmeier, Svendsen, and Vrchota (2004) reported a study in which rubrics were used for repeated teacher, self- and peer-assessments of oral communication skills in a food preparation course. The single group, longitudinal study required all the students to self- and peer assess a minimum of four oral presentations over the course of a semester. The average score of the presentations that semester was higher than the average score attained by students in a similar course the previous semester. In addition, the average grade for the course was 94%, as compared to 86% in the previous semester. The researchers also reported increases in interaction and participation in the classroom, along with enhanced understanding of food science. This latter deduction was made on the basis of judgments of more thoughtful and scientific questions asked towards the end of the semester. The authors attribute the development of higher-level thinking skills to the use of rubrics for self- and peer assessments.

A similar study which reported contrasting findings is by Green and Bowser (2006). Having used a rubric for evaluating the master's thesis literature reviews in a two-group post-test design ($N = 16$), a t -test conducted on the total score means showed that there were no significant differences between scores of samples written without and with rubrics. The contradictory findings can probably be explained in terms of the very small sample size, as well as the limited way in which the rubric was put to use in this study, which was to simply make the rubric available to the students prior to the submission of the reviews. Andrade (2001) has shown that just providing a rubric to middle school students is not consistently associated with better performance, and concludes that students must engage deeply with rubrics, perhaps by co-creating them and using them for self- and peer assessments, as students did in the Reitmeier, Svendsen, and Vrchota (2004) study.

Rubric use in instructional and programme assessments

While the studies discussed above examined the use of rubrics in course-based assessments, others explored their utility in terms of instructional and programme assessments. Dunbar, Brooks, and Kubicka-Miller (2006) documented a department's use of a rubric to ascertain instructional effectiveness in terms of oral communication skills. The study, conducted in seven classrooms of a general education public speaking course, involved the use of a rubric by two instructors to evaluate 100 student speeches at the end of a semester. Of the eight competencies delineated in the rubric, the students were rated 'unsatisfactory' on five. Having demonstrated the department's limited success in achieving its goals regarding public speaking skills, the authors provided feedback to the department. They suggested a re-examination of both the content of the course and the way in which it is delivered.

Similarly, Knight (2006) reports the results of using a rubric to grade bibliographies compiled by students in regular, service and honors sections of first-year undergraduate research and writing courses. The level of information literacy achieved by each of the sections was assessed using the rubric, and the information gained was used to map any consistently significant differences in learning due to classroom environments. The results were used to identify areas in need of improvement and to make changes to the library tutorial.

Song (2006) and Powell (2001) also illustrated the usefulness of rubrics in assessing the effectiveness of courses. Song tested a model to assess teacher candidates' ($N = 282$) growth in teaching performance, as well as intellectual and ethical readiness through performance-based artefacts rather than through test scores. Rubrics were used

to score students' performances. An examination of the scores for individual criteria provided feedback to the evaluators about students' overall strengths and weaknesses, and enabled them to identify areas for improvement in the instruction. Likewise, Powell found that the use of rubrics to rate student work enabled an instructor to pinpoint the areas of weakness and thereby identify needed improvements in the instruction.

A study by Petkov and Petkova (2006) provides an approach to designing rubrics for use across diverse courses in a programme. Providing evidence of attainment of agreement on the criteria and the levels of quality in a rubric between instructors of different information system courses, the authors developed scoring rubrics for projects in a post-graduate level programme. This approach to the development of a generic scoring rubric for use across courses has implications for the programme assessment as it enables comparability across courses and semesters.

These studies lend support to the view that rubrics have the potential to act as 'instructional illuminators' (Popham 1997, 75). Although the authors do not use the term instructional illuminators, they do deliberate upon the value of rubrics in identifying the understandings and skills to be taught and learned, and of providing detailed, criteria-specific feedback to instructors and departments on which of those understandings and skills have been mastered by students and which have not. As a result, the rubrics informed the process of making improvements to courses and instructions. The key to this process, of course, is a clear, valid and reliable rubric, without which the method is useless at best and possibly even misleading. Only two of the studies discussed above reported on the validity or reliability of the rubrics used; such information would significantly increase the credibility of the results reported by other authors.

Validity and reliability

Though few of the studies reviewed above reported on the validity and reliability of the rubrics used, a separate but related body of literature places considerable importance on testing the quality of a rubric by determining if it measures what it is intended to measure (validity) and provides for consistency in scoring (reliability) (Moskal and Leydens 2000; see also Jonsson and Svingby 2007). Little attention has been paid to the validity of rubrics. Most of the work has been on reliability – a necessary but insufficient condition of validity.

Reliability of rubrics

The types of reliability that are most often considered in classroom assessment and in rubric development involve rater reliability (Moskal and Leydens 2000) which refers to the consistency of scores that are assigned by two independent raters (inter-rater reliability) and by the same rater at different points in time (intra-rater reliability). The literature most frequently recommends two approaches to inter-rater reliability: consensus and consistency. While consensus (agreement) measures if raters assign the same score, consistency provides a measure of correlation between the scores of raters (Fleenor, Fleenor, and Grossnickle 1996). Some studies use generalisability theory to compute measurement estimates. The belief is that a well-designed scoring rubric should ameliorate inconsistencies in the scoring process by minimising errors due to rater training, rater feedback and the clarity of descriptions of criteria.

Several studies have shown that rubrics can allow instructors and students to reliably assess performance. Describing the development and application of a rubric

for assessing student portfolios at a Canadian university, Simon and Forgette-Giroux (2001) had students ($N = 100$) in four graduate and undergraduate courses use the rubric to self-assess their portfolios for formative and summative purposes. The average percentage agreement (consensus) between the professor's assessment and students' self-assessment was 75%. Using both consensus and consistency approaches, Hafner and Hafner (2003) examined the reliability between peer-grading and instructor grades in an undergraduate evolutionary biology course ($N = 107$). They concluded that the instructor and student ratings were remarkably uniform. In the study by Dunbar, Brooks, and Kubicka-Miller (2006), two of the authors graded student speeches ($N = 100$) in a foundational general education course. Using Ebel's intra-class correlation (consistency), the study showed high inter-coder reliability for the entire scale (.96) and for each of the evaluation criteria (.82–.97). Taken together, these studies indicate that the use of rubrics can lead to a relatively common interpretation of student performance.

There is also ample evidence of disagreement between assessors using rubric-referenced marking schemes. Using Cohen's kappa statistic, Oakleaf (2006) focussed on analysing the consistency with which rubric scores are assigned by multiple raters in a specific information literacy skill area. Seventy-five artefacts were assessed by 25 raters categorised into five-member groups. While *consistent* scoring ($k = 0.41$) of artefacts of student learning could be achieved using rubrics, different rater groups in this study arrived at varying levels of consensus (complete agreement) within their groups much below the acceptable levels ($k = 0.70$ or above).

Examining the reliability of the rubric by comparing the scores given by the raters to the scores given by the researcher, the study reports considerable variation across raters in terms of consensus. Beyond its contributions to the library literature, Oakleaf's study suggests that the most appropriate approach to ensuring reliability depends, in part, on how the results of an assessment are used. When calculating a total score for a grade (i.e. for summative purposes), consistency estimates may be adequate. If, however, decisions are made at the criterion level (such as when providing feedback on whether or not a particular criterion has been met), consensus estimates may be needed.

Boulet et al. (2004) stress scorer training as the most important factor for achieving reliable and valid large scale assessments. The study was conducted in the context of changes made to a patient note (PN) scoring rubric used to assess medical graduate students' ability to summarise and synthesise data collected from patients. To investigate the psychometric adequacy of PN scores, 61,497 PN ratings from 6225 candidates were scored by 3 to 10 raters using a rubric. In the pilot study, there was 76% variance due to between rater disagreements. To combat this in the decision study, the raters were trained twice as long before they did independent scoring. As a result of this increased training, the variance between raters decreased to 12%. Generalisability theory, a method of computing measurement estimates, was used in a persons by raters nested in cases ($p \times [\text{raters: cases}]$) design to estimate variance components and to provide measures of the reproducibility of the PN scores. The study reports two of the most frequently used statistics in generalisability analyses, namely the generalisability coefficient (σ^2) denoting relative error variance (consistency) and dependability coefficient (F) denoting absolute error variance (consensus). The generalisability coefficient of 0.71 and dependability coefficient of 0.65 indicated moderate reliability.

Boulet et al. (2004) also suggest ways of establishing the validity of rubric scores by conducting correlational studies. The strength of relationships of PN scores with

several internal and external performance measures such as prior training, medium of instruction, exposure to interviewing patients and so on, showed that all the internal and external variables were moderately related to the PN scores.

Validity of rubrics

Of the four papers found that discuss the validity of the rubrics used in research, three focussed on the appropriateness of the language and content of a rubric for the population of students being assessed. The language used in rubrics is considered to be one of the most challenging aspects of its design (Tierney and Simon 2004; Moni, Beswick, and Moni 2005). As with any form of assessment, the clarity of the language in a rubric is a matter of validity because an ambiguous rubric cannot be accurately or consistently interpreted by instructors, students or scorers (Payne 2003). This point is reinforced by Moni, Beswick, and Moni (2005), who revised a rubric with a view to improve the assessment quality of physiological concepts using group-constructed concept maps in a dentistry course. Data on the interpretation of rubric criteria collected by way of faculty reflections and student survey responses formed the basis for revising the descriptions of criteria and for incorporating new criteria.

The appropriateness of the language and content of rubrics for particular populations has been explored in other studies as well. Lapsley and Moody (2007) describe the development of a rubric made appropriate for two sets of students in an online human resources course by way of incorporating elements from their motivational learning styles. The objectives of the study were derived from the potential role of rubrics in channeling students' motivation and effort towards enhancing performance. It is based on the premise that traditional students are more focussed on passing the course, leading them to put less effort into learning as compared to non-traditional students, who are motivated by the potential of higher pay or a better job.

The study, conducted in four classrooms with different compositions of non-traditional and traditional students ($N = 151$), was designed to test whether or not the same rubric could be used by both sets of students to improve their performance. The results suggested that the assessment of traditional students using a rubric developed initially for non-traditional students did not bring about the required effort and quality of responses. Based on an understanding of the differences in the two sets of students, a revised rubric was developed. The revisions included converting the score levels (i.e. 5, 10, 14, etc.) to a score range (10–14, 15–16, etc.) and supporting the description of performance levels with detailed examples of possible student responses. Upon testing the revised rubric in a classroom, the authors found that it could provide an adequate assessment for both sets of students.

In another study that revealed the importance of adapting a rubric to the characteristics of the population of students to be assessed, Green and Bowser (2006) demonstrated the techniques for transferring a rubric developed in one institution to a second institution. A rubric developed for master's thesis literature reviews at Shenandoah University (SU) was used without any modification in a similar programme at Best Practices University (BPU). Literature reviews by commencing master's of education students at BPU were scored twice by paired raters from BPU. The same samples were then scored by raters from SU. The study reports that the raters at BPU consistently scored the samples higher than the raters from SU. The differences in scores between the raters of the two institutions was attributed to the fact that the rubric was designed for students concluding their literature reviews and theses but was being applied to the

work of students who were just beginning the literature review process – a problem of validity. The rubric was subsequently modified for use at BPU.

Taken together, these studies of validity and reliability indicate, at a minimum, the importance of attention to course goals and programme design during the development and use of rubrics, as well as the need for rater training. There is an unfortunate irony in the fact that one subset of the literature on rubrics strongly emphasises this issue while another subset largely ignores them.

Summary of review

This summary is framed in terms of the four questions that motivated the review.

In general, what kind of research has been done on rubrics in higher education?

Studies of rubrics in higher education have been undertaken in a wide range of disciplines in higher education, and for multiple purposes, including increasing student achievement, improving instruction and evaluating programmes. While the potential of rubrics to identify changes and improvements in course delivery and design has been demonstrated in four studies (Powell 2001; Dunbar, Brooks and Kubicka-Miller 2006; Knight 2006; Song 2006), the same amount of attention has not yet been paid to its usefulness in programme assessment. The limited research available, however, posits programme assessment as an area where rubrics can be used effectively (Petkov and Petkova 2006).

Studies of student and instructor perceptions have also been undertaken. While, student perceptions of rubrics are generally positive and some authors report positive responses to rubric use by instructors (Powell 2001; Reitmeier, Svendsen and Vrchota 2004; Andrade and Du 2005; Schneider 2006), at least two researchers have noted a tendency for instructors to resist using them (Bolton 2006; Parkes 2006). This resistance is probably due, at least in part, to the fact that the ‘overwhelming majority’ of instructors in higher education have little or no preparation as teachers, and minimal access to new trends in assessment (Hafner and Hafner 2003, 1510).

Hafner and Hafner also noted that rubric use could be limited by the perception that rubrics require a large investment of time and effort on the part of the instructor. We acknowledge that this perception is valid – why spend a lot of time figuring out a new way to do what we have done for decades? – but only when taking a narrow view of rubrics as scoring guides. We recommend educating instructors on the formative use of rubrics to promote learning by sharing or co-creating them with students in order to make the goals and qualities of an assignment transparent, and to have students use rubrics to guide peer and self-assessment and subsequent revision. When instructors appreciate the instructional leverage provided by student-involved rubric use, they are more likely to use one than when they see it only as a new but unnecessary approach to generating grades.

Is there evidence that rubrics can be used as formative assessments in order to promote learning and achievement in higher education, as opposed to rubric use that serves only the purposes of grading and accountability?

The results of the review in terms of this question were inconclusive. While the studies conducted by Petkov and Petkova (2006) and Reitmeier, Svendsen, and Vrchota

(2004) suggested that involving students in the development and use of rubrics or sharing an instructor-developed rubric prior to the submission of an assignment was associated with improvements in academic performance, Green and Bowser (2006) showed no differences in the quality of the work done by students with and without rubrics. The implication seems to be that simply handing out a rubric cannot be expected to have an impact on student work: students must be taught to actively use a rubric for self- and peer assessments and revision in order to reap its benefits. However, we hasten to note that the number of high quality studies that were found that address this question was quite small. More and more rigorous research on this issue is needed.

How much attention has been paid to the quality of the rubrics being used by college and university instructors?

A large majority of the studies reviewed did not describe the process of development of rubrics to establish their quality. The few studies that report inter-rater reliability (Simon and Forgette-Giroux 2001; Hafner and Hafner 2003; Dunbar, Brooks, and Kubicka-Miller 2006) tend to show that rubrics can lead to a relatively common interpretation of student performance. The important inference here is that raters must be sufficiently trained in order to achieve acceptable levels of reliability (typically 70% agreement or higher).

The research reports little study of the validity of the rubrics used. The few studies that do discuss it (Moni, Beswick, and Moni 2005; Green and Browser 2006; Lapsley and Moody 2007) have shown that clarity and appropriateness of language are central concerns. Important aspects of validity have not yet been addressed at all, including the need to establish the alignment between the criteria on the rubric and the content or subject being assessed (content validity); the facets of the intended construct being evaluated (construct validity); and the appropriateness of generalisations to other, related activities (criterion validity).

What are some fruitful directions for future research and development in this area?

We have identified four areas most in need of attention from the scholarly community: Rigorous research methodologies, geographical focus, validity and reliability, and the promotion of learning.

- (1) *More rigorous research methods and analyses.* More than half of the research reviewed for this paper did not utilise robust methodologies. Even those that did use experimental or quasi-experimental designs largely used post-test designs and did not control for important variables such as maturation, previous achievement, or the Hawthorne effect. In addition, most of the studies were conducted by the authors in their own classrooms. While this is an accepted approach, replications in similar or different contexts would strengthen the credibility of the results. Given the rather serious design limitations discussed above, we must consider the conclusions reported here with caution. More sophisticated research designs are needed in order to validate the preliminary findings. Time series designs and designs using pre- and post-tests are needed in order to establish the effectiveness of rubric-based interventions. We also

recommend the use of appropriate statistical tools such as ANCOVA to control for previous achievement, and the reporting of effect sizes.

- (2) *Expanded geographic and cultural perspectives.* This review revealed that research on this topic is limited almost entirely to the United States. The differences in educational theories and instructional approaches in different cultures necessitate international studies of rubric use in order to establish its utility in diverse contexts. Cross-cultural studies of student and instructor perceptions and usage of rubrics to understand the matches and mismatches between rubric-referenced assessment and learners' and teachers' conceptions of education are also needed.
- (3) *More research on validity and reliability.* The validity and reliability of the rubrics used in research have not received enough attention. Some studies mention having conducted pilot and reliability tests prior to the implementation of rubrics, however very few report the results. Information about the procedures, analyses and results would enable readers to better understand claims made about validity and reliability. Future studies should report how the validity of a rubric was established, and the scoring reliability, including rater training and its contribution towards achieving inter-rater reliability, and perhaps even the correlation between rubric-referenced scores and other measures of performance.
- (4) *A closer focus on learning.* A majority of the studies reported here illustrate the use of rubrics only for evaluation. The ways in which they can be used to teach has not been sufficiently addressed. Studies are needed that look beyond scores for an assignment by examining, for example, the development of positive attitudes towards and perceptions about learning, the acquisition and integration of new knowledge, extending and refining knowledge, using knowledge to perform meaningful tasks, and developing powerful habits of mind that enable students to regulate their behaviour and think critically and creatively (Marzano et al. 1988).

In a similar vein, though studies of interventions that involve simply handing out rubrics have value, more can be done in this area in post-secondary contexts. While some preliminary work has been done on peer and self-assessment (Andrade 2001; Reitmeir, Svendsen, and Vrchota 2004), research on the relationships between rubrics and self-regulatory behaviour in students in higher education would be illuminating.

Acknowledgement

The authors are grateful to Dr Anders Jönsson for his thoughtful feedback on a draft of this manuscript.

Notes on contributors

Y. Malini Reddy is an assistant professor in the discipline of marketing at ICAI Business School, Hyderabad, India. She is pursuing her doctoral research on the relationships between rubrics and student learning.

Heidi Andrade is an assistant professor in educational psychology and methodology at the University at Albany in Albany, New York, USA. Her research focuses on classroom assessment, with a focus on student self-assessment and self-regulation.

References

- Andrade, H. 2000. Using rubrics to promote thinking and learning. *Educational Leadership* 57, no. 5: 13–18.
- Andrade, H.G. 2001. The effects of instructional rubrics on learning to write. *Current Issues in Education* 4, no. 4. <http://cie.ed.asu.edu/volume4/number4> (accessed January 4, 2007).
- Andrade, H.G. 2005. Teaching with rubrics: The good, the bad, and the ugly. *College Teaching* 53, no. 1: 27–30.
- Andrade, H., and Y. Du. 2005. Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation* 10, no. 5: 1–11.
- Andrade, H., Y. Du, and X. Wang. 2008. Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practices* 27, no. 2: 3–13.
- Arter, J., and J. Chappuis. 2007. *Creating and recognizing quality rubrics*. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- Arter, J., and J. McTighe. 2001. *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance*. Thousand Oaks, CA: Corwin/Sage.
- Bolton, C.F. 2006. Rubrics and adult learners: Andragogy and assessment. *Assessment Update* 18, no. 3: 5–6.
- Boulet, J.R., T.A. Rebbecki, E.C. Denton, D. McKinley, and G.P. Whelan. 2004. Assessing the written communication skills of medical school graduates. *Advances in Health Sciences Education* 9: 47–60.
- Campbell, A. 2005. Application of ICT and rubrics to the assessment process where professional judgment is involved: the features of an e-marking tool. *Assessment & Evaluation in Higher Education* 30, no. 5: 529–37.
- Cohen, E.G., R.A. Lotan, P.L. Abram, B.A. Scarloss, and S.E. Schultz. 2002. Can groups learn? *Teachers College Record* 104, no. 6: 1045–68.
- Dunbar, N.E., C.F. Brooks, and T. Kubicka-Miller. 2006. Oral communication skills in higher education: Using a performance-based evaluation rubric to assess communication skills. *Innovative Higher Education* 31, no. 2: 115–28.
- Fleenor, J.W., J.B. Fleenor, and W.F. Grossnickle. 1996. Interrater reliability and agreement of performance ratings: A methodological comparison. *Journal of Business and Psychology* 10: 367–80.
- Green, R., and M. Bowser. 2006. Observations from the field: Sharing a literature review rubric. *Journal of Library Administration* 45, nos. 1–2: 185–202.
- Hafner, J.C., and P.M. Hafner. 2003. Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education* 25, no. 12: 1509–28.
- Jonsson, A., and G. Svingby. 2007. The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review* 2: 130–44.
- Knight, L.A. 2006. Using rubrics to assess information literacy. *Reference Services Review* 34, no. 1: 43–55.
- Lapsley, R., and R. Moody. 2007. Teaching tip: Structuring a rubric for online course discussions to assess both traditional and non-traditional students. *Journal of American Academy of Business* 12, no. 1: 167–72.
- Marzano, R.J., R.S. Brandt, C.S. Hughes, B.F. Jones, B.Z. Presseisen, C. Stuart, and C. Shor. 1998. *Dimensions of thinking*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Moni, R.W., E. Beswick, and K.B. Moni. 2005. Using student feedback to construct an assessment rubric for a concept map in physiology. *Advances in Physiology Education* 29: 197–203.
- Moskal, B.M., and J.A. Leydens. 2000. Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation* 7, no. 10: 71–81. <http://pareonline.net/getvn.asp?v=7&n=10> (accessed July 9, 2007).
- Oakleaf, M.J. 2006. Assessing information literacy skills: A rubric approach. PhD diss., University of North Carolina, Chapel Hill. UMI No. 3207346.
- Osana, H.P., and J.R. Seymour. 2004. Critical thinking in preservice teachers: A rubric for evaluating argumentation and statistical reasoning. *Educational Research and Evaluation* 10, nos. 4–6: 473–98.

- Parke, C.S. 2001. An approach that examines sources of misfit to improve performance assessment items and rubrics. *Educational Assessment* 7, no. 3: 201–25.
- Parkes, K.A. 2006. The effect of performance rubrics on college level applied studio grading. PhD diss., University of Miami. UMI No. 3215237.
- Payne, D.A. 2003. *Applied educational assessment*. 2nd ed. Belmont, CA: Wadsworth/Thomson Learning.
- Petkov, D., and O. Petkova. 2006. Development of Scoring Rubrics for IS Projects as an Assessment Tool. *Issues in Informing Science and Information Technology* 3: 499–510.
- Popham, W.J. 1997. What's wrong – and what's right – with rubrics. *Educational Leadership* 55, no. 2: 72–5.
- Powell, T.A. 2001. Improving assessment and evaluation methods in film and television production courses. PhD diss., Capella University. UMI No. 3034481.
- Quinlan, A.M. 2006. *A complete guide to rubrics, assessments made easy for teachers, K-College*. USA: Rowman & Littlefield Education.
- Reitmeier, C.A., L.K. Svendsen, and D.A. Vrchota. 2004. Improving oral communication skills of students in food science courses. *Journal of Food Science Education* 3: 15–20.
- Schneider, J.F. 2006. Rubrics for teacher education in community college. *The Community College Enterprise* 12, no. 1: 39–55.
- Simon, M., and R. Forgette-Giroux. 2001. A rubric for scoring postsecondary academic skills. *Practical Assessment, Research and Evaluation* 7, no. 18. Available <http://pareonline.net/getvn.asp?v=7&n=18> (accessed December 7, 2006).
- Song, K.H. 2006. A conceptual model of assessing teaching performance and intellectual development of teacher candidates: A pilot study in the US. *Teaching in Higher Education* 11, no. 2: 175–90.
- Stiggins, R.J. 2001. *Student-involved classroom assessment*. 3rd ed. Upper Saddle River, NJ: Prentice-Hall.
- Tierney, R., and M. Simon. 2004. What's still wrong with rubrics: focusing on the consistency of performance criteria across scale levels'. *Practical Assessment, Research & Evaluation* 9, no. 2.
- Tunon, J., and B. Brydges. 2006. A study on using rubrics and citation analysis to measure the quality of doctoral dissertation reference lists from traditional and nontraditional institutions. *Journal of Library Administration* 45, nos. 3–4: 459–81.

Copyright of Assessment & Evaluation in Higher Education is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.