Rubrics
Annotated Bibliography (regularly updated)

Notes: Provides a definition, an example, reasons for use, and instructions for creating rubrics. The concluding paragraph marks Andrade’s transition into studying the self-assessment.


Abstract: This article gives a brief overview of the structure and purposes of rubrics; reviews the benefits of using rubrics as both teaching and grading tools; warns against approaches that limit the effectiveness of rubrics; and urges instructors to take simple steps toward ensuring the validity, reliability, and fairness of their rubrics. Tips for using rubrics with undergraduate and graduate students are also included.


Abstract: The authors examined the impact of self-assessment on 7th- and 8th-grade students’ written essays. Students wrote 2 essays: historical fiction essay and response to literature essay. All students received instructional rubrics that articulated the criteria and gradations of quality for the given essay. Students in the treatment group participated in 2 formal self-assessment lessons, during which they used the rubric to assess the quality of their drafts. Authors used multiple linear regression to examine the relationship between essay scores, treatment, and a set of control predictors. The results from the historical fiction essay suggested a positive relationship between the treatment and girls’ scores, but no statistically significant relationship between the treatment and boys’ scores. The results from the response to literature essay show no effect of treatment for either boys or girls. The results are explained in terms of the insufficiency of the intervention, as well as the possible effects of rubrics, school conditions, and gender differences in response to self-generated feedback.

Abstract: This study suggests that students use rubrics to support their own learning and academic performance. In focus groups, fourteen undergraduate students discussed the ways in which they used rubrics to plan an approach to an assignment, check their work, and guide or reflect on feedback from others. The students said that using rubrics helped them focus their efforts, produce work of higher quality, earn a better grade, and feel less anxious about an assignment. Their comments also revealed that most of the students tend not to read a rubric in its entirety, and that some may perceive of a rubric as a tool for satisfying a
particular teacher’s demands rather than as a representation of the criteria and standards of a discipline.


Abstract: This study investigated the relationship between middle school students' scores for a written assignment (N = 162) and a process that involved students in generating criteria and self-assessing with a rubric. Gender, time spent writing, grade level, prior rubric use, and previous achievement in English were also examined. The treatment involved using a model essay to scaffold the process of generating a list of criteria for an effective essay, reviewing a written rubric, and using the rubric to self-assess first drafts. The comparison condition involved generating a list of criteria and reviewing first drafts. Findings include a main effect of treatment, gender, grade level, writing time, and previous achievement on total essay scores, as well as main effects on scores for every criterion on the scoring rubric. The results suggested that reading a model, generating criteria, and using a rubric to self-assess can help middle school students produce more effective writing.


Abstract: The purpose of this study was to investigate the effect of reading a model written assignment, generating a list of criteria for the assignment, and self-assessing according to a rubric, as well as gender, time spent writing, prior rubric use, and previous achievement on elementary school students' scores for a written assignment (N = 116). Participants were in grades 3 and 4. The treatment involved using a model paper to scaffold the process of generating a list of criteria for an effective story or essay, receiving a written rubric, and using the rubric to self-assess first drafts. The comparison condition involved generating a list of criteria for an effective story or essay, and reviewing first drafts. Findings include a main effect of treatment and of previous achievement on total writing scores, as well as main effects on scores for the individual criteria on the rubric. The results suggest that using a model to generate criteria for an assignment and using a rubric for self-assessment can help elementary school students produce more effective writing.


Abstract: The authors investigated the relation between long- and short-term rubric use (including self-assessment), gender, and self-efficacy for writing by elementary and middle school students (N = 268). They measured long-term rubric use with a questionnaire. They manipulated short-term rubric use by a treatment that involved reviewing a model and using a rubric to self-assess drafts. The authors collected self-efficacy ratings 3 times. Results revealed that girls' self-efficacy was higher than boys' self-efficacy before they began writing. The authors found interactions between gender and rubric use: Average self-efficacy ratings increased as students wrote, regardless of condition, but the increase in the self-efficacy of girls in the treatment group was larger than that for girls in the comparison group, and long-term rubric use associated only with the self-efficacy of girls.


Abstract: The article discusses feedback in education, and offers advice for delivering effective feedback. The assertion that feedback must be timely, specific, positive, and tailored to the specific students' needs and understanding is examined, and the advice to ensure that feedback is descriptive of the work rather than the student is presented. The argument that feedback is not useful to students if they do not have timely opportunities to apply it and if they do not have a clear, specific goal to work towards is also commented on.


Abstract: The United States education system depends on legislation and funding at the federal, state and local levels. Public understanding of assessment therefore is important to educational reform in the USA. Educational reformers often invoke assessment information as a reason for reform, typically by citing unacceptable achievement on some measure or indicator. Recent educational reforms in the US also rely on assessment information as evidence of the effectiveness of the reform, designing some sort of accountability system into the reform. Public opinion about testing in three recent waves of US educational reform (the minimum competency movement in the 1970s, the standards-based reform movement in the 1980s and 1990s, and the No Child Left Behind era beginning in 2002) shows two themes. One is public belief in the objectivity of testing. The other is public belief in using tests data comparatively and competitively.


Abstract: The authors investigated the relationship between scaffold types [rubrics] and the level of students’ thinking skill performance, learning achievement and attitudes, in a two-by-two factorial, quasi-experimental study. Participants included 216 undergraduate preservice K-12 teachers who were presented with one of four versions of an asynchronous discussion board assignment. Resulting discussion interactions were evaluated for demonstration of low-, mid- and higher-order thinking skills. Findings revealed students who were given a scaffold demonstrated higher-level thinking skills more frequently than did students who received no scaffold. No significant differences in learning achievement associated with test performance were found in test results. The treatment variables did significantly affect effect survey ratings associated with students’ attitudes.


Abstract: Although the rubric has emerged as one of the most popular assessment tools in progressive educational programs, there is an unfortunate dearth of information in the literature quantifying the actual effectiveness of the rubric as an assessment tool in the hands of the students. This study focuses on the validity and reliability of the rubric as an assessment tool for student peer-group evaluation in an effort to further explore the use and effectiveness of the rubric. A total of 1577 peer-group ratings using a rubric for an oral presentation was used in this 3-year study involving 107 college biology students. A quantitative analysis of the rubric used in this study shows that it is used consistently by both students and the instructor across the study years. Moreover, the rubric appears to be 'gender neutral' and the students’ academic strength has no significant bearing on the way that they employ the rubric. A significant, one-to-one relationship (slope = 1.0) between the instructor's assessment and the students' rating is seen across all years using the rubric. A generalizability study yields estimates of inter-rater reliability of moderate values across all years and allows for the estimation of variance components. Taken together, these data indicate that the general form and evaluative criteria of the rubric are clear and that the rubric is a useful assessment tool for peer-group (and self-) assessment by students. To our knowledge, these data provide the first statistical documentation of the validity and reliability of the rubric for student peer-group assessment.


Abstract: The purpose of the study was to offer some quantitative, multivariate evidence concerning the impact of grading rubric use on academic outcome among American higher education students. Using a pre-post, quasi-experimental research design, cross-sectional data were derived from undergraduates enrolled in an elective during spring and fall 2009 at an American research university. The control group, who completed an assignment without a grading rubric, scored significantly lower, on average, than the treatment group who completed the same assignment with the use of the rubric. Regression results indicated that the grading rubric constituted a significant predictor of assignment outcome, but exerted a less powerful effect than baseline course knowledge and grade point average. Suggestions for future research are provided.


Abstract: Although feedback has a great potential for learning, students do not always make use of this potential. This article therefore reviews research literature on students’ use of feedback in higher education. This is done in order to find answers as to why some students do not use the feedback they receive and which factors are important in influencing students’ use of teacher feedback.
Findings show that utility is not only a key feature for students’ use of feedback but also that some factors, such as lack of strategies for productively using feedback or lack of understanding of academic discourse, may hinder students’ possibilities to use the information formatively.


Abstract: This paper reports on a study where rubrics have been used to convey assessment expectations to students (n = 176) in three different assessment situations in professional education. These situations are: (1) the development of a survey instrument, which was part of a course in statistics and epidemiology; (2) an inspection of a house, which was part of a course about the functions of buildings for real estate brokers and (3) a workshop in communication with patients, which was part of a course in the evaluation of diagnostic procedures and treatments of oral infections in dental education. In all situations, students’ perceptions and uses of the rubrics were investigated. Findings suggest that it is indeed possible to convey expectations to students through the use of rubrics, in the sense that students not only appreciate the efforts to make assessment criteria transparent, but may also use the criteria in order to support and self-assess their performance. Important features of the rubrics, which were found to facilitate students’ understanding and use of the criteria in these situations, are presented and discussed.


Abstract

Several benefits of using scoring rubrics in performance assessments have been proposed, such as increased consistency of scoring, the possibility to facilitate valid judgment of complex competencies, and promotion of learning. This paper investigates whether evidence for these claims can be found in the research literature. Several databases were searched for empirical research on rubrics, resulting in a total of 75 studies relevant for this review. Conclusions are that: (1) the reliable scoring of performance assessments can be enhanced by the use of rubrics, especially if they are analytic, topic-specific, and complemented with exemplars and/or rater training; (2) rubrics do not facilitate valid judgment of performance assessments per se. However, valid assessment could be facilitated by using a more comprehensive framework of validity when validating the rubric; (3) rubrics seem to have the potential of promoting learning and/or improve instruction. The main reason for this potential lies in the fact that rubrics make expectations and criteria explicit, which also facilitates feedback and self-assessment.


Abstract In this study, the primary school teachers’ attitudes towards rubrics were investigated with reference to different variables that are related to their use of these tools. In this survey model study, the data were analyzed through one-way and two-way ANOVA tests. The results of the study indicate that the teachers who consider their level of information on rubrics quite sufficient and those who use the rubrics in order to monitor the development of students’ advanced thinking skills as well as in order to give feedback on student tasks have higher attitude scores. Furthermore, the results also show that the teachers with higher number of students and those who use rubrics in order to grade students have lower attitude scores.

**Abstract**

The purpose of this study was to examine the views of primary school teachers with negative and positive attitudes towards rubrics comparatively in terms of various variables. 292 teachers from 17 primary schools in Ankara province of Turkey participated into the study. The data obtained from the views of teachers with positive and negative attitudes towards rubrics were examined with figures. The findings indicated that the teachers with positive attitudes towards the rubrics benefited from the rubrics more than the teachers with negative attitudes when they were evaluating the achievement of the students in performance tasks. Both the teachers with positive and negative attitudes stated that they had knowledge about the rubrics and they used the course books while getting this information. Moreover, the teachers with negative attitudes consulted their colleagues more than the teachers with positive attitudes. It was observed that the teachers with positive attitudes prepared the rubrics themselves more often than the teachers with negative attitudes and they used the information obtained from the rubrics for “observing the development of students’ higher order thinking skills” and “observing to what extent the students used their knowledge and skills in real life situations”. The teachers with negative attitudes used the information obtained from the rubrics for “grading”.


Notes: Refutation of the arguments and the logic of the arguments against the use of rubrics. The “struggle against inappropriate use of rubrics by lazy instructors or lazier administrative bureaucrats of one stripe or another is commendable” p. 109.


**Abstract:** Research on formative classroom assessment practices and professional development is richer at the classroom/teacher level than at the building/administrator level. Yet administrator leadership is known to be critical for school reforms, including a change to more formative, learning-oriented assessment practices. The researchers conducted an exploratory study using two years of data from a large, rural school district to describe administrators' learning as they participated in a professional development project designed to increase their knowledge and leadership of formative assessment. Teachers skilled at formative assessment, in the view of their administrators, shared learning targets with students in multiple ways and before, during, and after the lesson. The leadership of administrators was critical to the implementation of formative classroom assessment in their buildings. Administrators who became the leading learner focused their observations in classrooms more intently than before on what the students (not just the teachers) were doing. They understood formative assessment themselves, and their schools were the ones that made progress implementing formative assessment.


Notes (my language): Love the quote: “the vast majority of rubrics are instructionally fraudulent” p. 73. Rubrics should focus attention on process over results. Knowing task specific particulars seldom has value to skill development, synthesis, or practical applications. The creation of amorphous levels of achievement result in reification of initial molar false impressions that only appear to be discerning judgments. That is, there is nothing formative for the student on how to improve on the skill. Poor rubrics require too much. Such an error probably indicates a lack of appreciation for the steps in the process of developing and refining the skill. Put another way, putting too many criteria and levels on a rubric reveals the professor’s focus on summative evaluation as opposed to formative feedback. A key test of a rubric is if users can apply it to multiple assignments or works that represent the expected skill development. If there is only one possible assignment application for the rubric, it is not actually a rubric. Further, such a situation reveals a likely flaw in the instructional design for the course.


Abstract:
The mainstream research on scoring rubrics has emphasized the summative aspect of assessment. In recent years, the use of rubrics for formative purposes has gained more attention. This research has, however, not been conclusive. The aim of this study is therefore to review the research on formative use of rubrics, in order to investigate if, and how, rubrics have an impact on student learning. In total, 21 studies about rubrics were analyzed through content analysis. Sample, subject/task, design, procedure, and findings, were compared among the different studies in relation to effects on student performance and self-regulation. Findings indicate that rubrics may have the potential to influence students learning positively, but also that there are several different ways for the use of rubrics to mediate improved performance and self-regulation. There are a number of factors identified that may moderate the effects of using rubrics formatively, as well as factors that need further investigation.


Abstract: The objective of this study was to compare the effects of situations in which self-assessment was conducted using rubrics and situations in which no specific self-assessment tool was used. Two hundred eighteen third year pre-service teachers were assigned to either non-rubric or rubric self-assessment for designing a conceptual map. They then assessed their own maps. The dependent variables were self-regulation measured through a questionnaire and an open question on learning strategies use, performance based on an expert-assigned score, accuracy comparing self-scores with the expert’s scores and task stress using one self-reported item. The results showed that the rubric group reported higher learning strategies use, performance and accuracy. However, the rubric group also reported more problems coping with stress and higher performance/avoidance self-regulation that was detrimental to learning.

Abstract: Construct validity of peer assessment (PA) is important for PA application, yet difficult to achieve. The present study investigated the impact of an assessment rubric and friendship between the assessor and assessee on construct validity of PA. Two-hundred nine bachelor students participated: half of them assessed a peer’s concept map with a rubric where as the other half did not use a rubric. The results revealed a substantial reliability and construct validity for PA. All students over-score their peers’ performance, but students using a rubric were more valid. Moreover, when using a rubric a high level of friendship between assessor and assessee resulted in more over-scoring. Use of a rubric resulted in higher quality concept maps for peer and expert ratings.


Abstract

Purpose - This paper seeks to discuss the characteristics that describe a rubric. It aims to propose a systematic method for developing curriculum wide rubrics and to discuss their potential utility for program quality assessment.

Design/methodology/approach - Implementation of rubrics is a recent phenomenon in higher education. Prior research and theoretical issues related to rubric design and development are discussed. The proposed method for rubric development is illustrated by deriving generic analytic scoring rubrics for two assessment methods, namely projects and cases in a Master's level business program in India. Aspects related to the validity of the rubrics developed are investigated and results of reliability study conducted using intraclass correlation coefficient (ICC) are reported. Findings - Upon testing, the rubrics were found to be reliable and valid grading tools. Results of inter-rater and intra-rater reliability analyses illustrated that the rubrics supported standardization of the manner in which cases and projects could be evaluated in different business courses.

Practical implications - Whereas previous studies on rubric construction have largely concentrated on task specific rubrics, this study focuses on development of curriculum wide rubrics that can be employed for assessment of students' learning at both course and program level. Originality/value - To date there has not been any published work on issues of assessment of student learning through project and case analysis rubrics within diverse courses in a business program. The method detailed in the study can guide the development of generic rubrics for alternative assessment methods employed in business programs as well as in other disciplines.


Abstract: 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.


Abstract: The 2001 U.S. Supreme Court Case of Falvo v. Owasso School System (Owasso Independent School District No I-011 v. Falvo) has focused national attention on the common classroom practice of peer-grading. In a unanimous decision the court reaffirmed the popular view that students grading each others’ tests is valuable, saving teachers’ time and augmenting student learning. Our study puts these presumed benefits to the test in 4 middle school science classrooms. We compared teacher-assigned grades to those awarded either by students to themselves or by their peers. By training students to grade with the help of a scoring rubric, a very high correlation was obtained between students and their teacher on test questions (r = 0.91 to 0.94). We found patterns of bias when students assigned grades. When grading others, students awarded lower grades to the best performing students than their teacher did. When grading themselves, lower performing students tended to inflate their own low scores. Performance on an unannounced, 2nd administration of the same test 1 week later measured the degree to which student-grading resulted in any increased understanding. Students who graded their peers’ tests did not gain significantly more than a control group of students who did not correct any papers but simply took the same test again. Those students who corrected their own tests improved dramatically. Self-grading and peer-grading appear to be reasonable aids to saving teachers’ time. Self-grading appears to result in increased student learning; peer-grading does not.


Abstract: When assessment tasks are set for students in universities and colleges, a common practice is to advise them of the criteria that will be used for grading their responses. Various schemes for using multiple criteria have been widely advocated in the literature. Each scheme is designed to offer clear benefits for students. Breaking down holistic judgments into more manageable parts is seen as a way to increase openness for students and achieve more objectivity in grading. However, such approaches do not adequately represent the full complexity of multi-criterion qualitative judgments, and can lead to distorted grading decisions. Six anomalies in the ways assessors approach the grading task are identified, together with several likely contributing factors. Overall, the conclusion is that explicit grading models do not have as strong a theoretical foundation as is commonly supposed, and that holistic appraisal merits further investigation.

Notes: Appears to reject rubrics primarily because many are poorly developed and poorly used. He complains strongly about the emphases on reliability at the expense of validity. “Limiting...to preset criteria [evaluators...can...take into account all the necessary nuances of expert judgments” (p. 20). Observations: 1. Scoring does not compute into an overall grade. Raters blur judgments across criteria. Raters make holistic judgments and then make the scores on the criteria fit with that judgment; 2. Overall evaluations do not match the scores on the criteria; 3. Missing criteria often matter most. It is the innovation that often wins the prize; 4. Criteria are mutually exclusive (or “distinct”); 5. Different rubrics for the same type of assignment will often have different criteria. Any evaluation, therefore will necessarily be missing relevant criteria; 6. While the overall scores are consistent, there are differences between assessors on the criteria scores.

Abstract: For students to develop the capacity to produce complex works of consistently high quality, they need to be able to monitor the quality of their work during its production. This requires that they know what constitutes high quality; how to compare the quality of their emerging work with their model of high quality; and how to modify their work accordingly. A common approach to assessing complex works is to apply a set of fixed criteria. This practice gives rise to various anomalies, two of which are analysed in this chapter. This phenomenon calls into question the fundamental validity of using preset criteria as the preferred approach for judging quality. Instead, holistic judgments are required. A teaching approach that deliberately blurs the boundary between pedagogy and assessment so students can develop the ability to make holistic appraisals is outlined in the second part of this chapter.


Notes: Traces the evolution of assessment thinking to argue that reliance on standardized testing is generally outdated and interferes with effective educational practices. Instead, teachers should be required to demonstrate competence with in-assessment methods. She notes that the UK made a transition away from the reliance on standardized testing in the 1990s. “[End-of-year tests] are like shopping mall medical screenings” (p. 13).


Abstract: Assessment tasks can focus student efforts in what they choose to study. This paper describes the development and initial application of an assessment matrix to support student learning in postgraduate research studies at the University of Adelaide. Students and supervisors in the multi-disciplinary Master programmes also benefit from the clear descriptors to clarify expectations and standardise grading. This reflective paper explores the capacity of a matrix to improve constructive alignment between research course objectives and learning outcomes when it is an integral part of feedback loops.
