Scoring Rubrics – An Annotated Bibliography/Review of Selected Publications

Thomas W. Zane PhD ~ Last Edited in October, 2012.

Introductory Notes

This bibliography/review contains my thoughts on many recent articles and books related to the construction and use of scoring rubrics in education. Each annotation/review offers a thumbnail sketch of findings or other information that may be of use to rubric developers and/or users. In several instances, I break the fourth wall to describe personal asides and experiences. Although this work is designed for academics, no attempt was made to remain objective or neutral in each annotation.

I have provided links for accessing many of the listed works when those articles and books are freely available online. Most of the unlinked articles in this bibliography are available via the Educational Resources Information Center (ERIC) and most of the monographs are available through academic interlibrary loan.

Disclaimer: This is an unedited, personally biased, “stream of consciousness” document designed to foster discussion about the materials listed. It is not intended to be a fair and balanced review of the literature.

Navigation: Turn on the Microsoft Word Navigation Pane (View/Navigation Pane) to see all the titles in this bibliography.

Please e-mail me with suggestions for improvement and counter points to my comments.


In this article the authors attempt to move teachers’ thinking beyond measuring the body of knowledge. Their rubrics are adequate but would need work to bring them up to standard. They do offer some ideas that might be useful for measuring students' success in open-ended, group, and problem-solving activities. They close with a series of real-world questions that teachers could ask themselves to help them build their own rubrics.

The audience for this article is any faculty member looking to build problem-based rubrics. However, note that the examples the authors provide are not the best available in the literature.

Rubric Examples: Several – all designed around classroom activities and projects. None of the examples are very strong.


This article is one of the author’s earlier works dealing with rubrics. She presents many good examples of analytic rubrics. This is important for two reasons. First, few authors—especially at
this early date—present immediately usable rubrics. Andrede’s are quite good. Second, because at the time of writing this piece, much of the academy was still using holistic rubrics. The author offers what I assume are anecdotally based speculations suggesting that rubrics are most effective when the language and standards are very clear and when students are involved in making the rubrics. Later works from this same author show evidence of several of these claims. This article also offers one of the earlier step-by-step approaches to building rubrics.

The audience for this article is any faculty member. This is a solid article from one of the most prolific scholars studying rubrics in the past 20 years. Readers may also be interested in later works from this same author.

Rubric Examples: Many examples of rubrics. Highly recommended.


The author describes what rubrics are, what they are good for, and points out that rubrics need to be very clear to all stakeholders to be effective. She argues that students could be involved with building the rubrics that will be used to score student work, or at least be included in a classroom discussion about those rubrics before using them. The idea is that getting students this involved will maximize the usefulness of the rubrics for supporting learning.

Of particular note is the author’s contention that rubrics can be a strong component of learning and assessment. This contention is supported in other literature, in that students who receive rubrics tend to do better on assignments, and teachers who pay the price to build rubrics seem to do a better job of conceptualizing what they expect from their students.

The audience for this article is any group of educators, teachers, and academic leaders who might need to use rubrics in their classrooms. The article offers a good basic—albeit brief—introduction to rubrics.

Rubric Examples – Reports, general writing, oral presentation.


Like the previous article, the author does a good job of covering what rubrics are, what they are good for, and how you might use them. But in this article she spends more time discussing validity, reliability, and fairness.

The author admits that this article is based on reflections about her use of rubrics in the classroom. The author does a better job (than in her previous articles) of outlining exactly what makes for good (and bad) rubrics – which gives the reader a fuller view of rubrics than with her two previous articles. She also does a better job of making a case for aligning rubrics to standards.

The audience for this article is any higher education faculty members.
Rubric Examples – Reports, general writing, oral presentation. I question the validity of the “collaboration” measure—as graders would not have a good opportunity to see the stated behaviors. The rubric in Table 1 contains nearly holistic measures on each row.


This article clearly builds on Andrade’s previous articles and work. She explores how rubrics can support learning via self assessment. She believes that students can use rubrics to see what it is they need to do, and then determine for themselves how well they are performing. As with her previous articles, her comments seem grounded in the literature, anecdotal evidence from her own classroom experiences, and some minor data collection. Even so, (as she did in previously published articles) she makes compelling arguments.

The author opened my eyes to additional benefits of using rubrics for formative evaluation practice. She offered a few rules that should maximize those benefits. First, be sure to give students the rubrics early enough to aid learning and formative self assessment. Second, don’t entangle this formative assessment with grading. Third, it is possible to create rubrics that are so tightly defined that they are really just prescriptions for the assignment – thus destroying any creativity. Students end up simply parroting exactly what the rubrics say. This is a critical issue that is not well explored in other rubric literature. Therefore, be careful to write the rubrics at just the right level of detail to foster success.

The audience for this article is any teacher.

Rubric Examples: None.


This article is a research study that follows up on the previous articles from Heidi Andrade. It is perhaps her first foray into a more experimental approach to studying rubrics. In this article the authors find evidence for Andrade’s speculative comments in her 2007 article (reviewed previously) namely, that giving students rubrics ahead of time improves their performance.

The audience for this document is not clear. Given its format, tone, and publisher, it seems to be aimed at those who deal with developing rubrics at a professional level. However, the document is quite readable and may be useful to anyone researching rubrics.

Rubric Examples: Fourth-grade writing assignment. The rubric has some errors. For example, “organization” adds aspects to differentiate a high score.

This is another in a series of articles from Heidi Andrade. She and her graduate students begin with a brief literature review that illustrates what I have mentioned elsewhere in this bibliography. Many authors present speculative lists of benefits of rubrics and rubric use, they cite one another, or they offer anecdotal evidence of the benefits. Few studies offer empirical evidence for or against the posited benefits. In this study, the authors searched for evidence of the linkage between rubric use and student self-efficacy (student beliefs about their ability to learn and achieve goals) and extended links to learning and achievement.

The study findings revealed positive relationships between rubric use and self-efficacy. The authors speculate that the underlying linkage relates to Bandura’s theory dealing with mastery experiences. As students use the rubrics and see successful outcomes, then self-efficacy improves. Regardless of the underlying cause, the increases in self-efficacy are encouraging.

The audience for this article is anyone studying rubrics. The article is very readable and provides a solid skeptical stance I appreciate in scientific papers.

Rubric Examples: Very solid writing rubrics designed to be teaching and self-evaluation tools. These rubrics contain good examples of meaningful concrete descriptors across the rows.


This work presents one of the most detailed descriptions of how to develop rubrics. The authors “show to use clear criteria and rubrics to make fair, consistent judgments and how rubrics can improve student work. The authors cover many specific issues with building better rubrics. They begin with a discussion of the nature, purposes, and benefits of using solid performance criteria. They explain the various types of rubrics in chapter two. Perhaps because of the relatively early publication date, this chapter is a bit brief and leaves out several other types of rubrics discussed elsewhere in the literature. Chapter 3 is dedicated to designing and developing rubrics. While a useful process, I felt this chapter was not the authors’ strongest. Other authors do a better job of outlining and detailing steps for successful rubric development (i.e., Quellmalz, 1991; Relearning by Design, 2000; and Rhodes, 2010). Chapter four provides an adequate metarubric (a rubric for judging the quality of your rubric) and Chapter five discusses converting rubric scores into grades. The last chapter was quite different for a text aimed at developing rubrics. The authors go into detail about how to use rubrics for learning and teaching. They even provide several strategies for using rubrics in different situations. This chapter is worth reading because other authors do not provide this sort of detailed analysis of uses for rubrics.

According to series editors the audience is “teachers, school leaders, policy makers, government officials, and those concerned with these crucial aspects of educational reform.” I question this broad statement. The textbook provides step-by-step descriptions of disentangling standards from criteria and then how to turn those into good performance measures. I can’t see anyone but educators reading this book. What is disappointing about this text is the lack of good examples.
Rubric Examples: Many. I caution the reader, some of the rubrics are quite good while others need major work to be useful. Still others are meant as bad examples that the authors use to make their points about what can go wrong with rubrics.

Metarubric (a rubric used to score the quality of rubrics).


This article is perhaps one of the very best descriptions of the logic and reasoning needed to take a nebulous concept like “science knowledge” and break it down into measures. However, a word of caution before I launch. While their process was perhaps the best I’ve seen, the resulting rubric was substandard.

The authors begin by designing a measurement model – what they call a “continuum for assessing science process knowledge” (p. 3). Then they outline the potential elements within this continuum by operationalizing science-based constructs and a taxonomy of learning. The result is a series of what they call “rationale” (p. 4) which justifies the operationalized constructs for measurement. But then they return to their measurement model (as they should) and map their constructs into levels along the continuum (i.e., Emerging, Beginning, Developing, Advancing, and Consolidating). Finally, the authors work to improve their model after pilot testing.

The nature of what the teachers did, and the ways the authors discuss the work is enlightening. So don’t let the narrative descriptive format fool you. This article presents a good look at a strong process for developing a measurement model before building a rubric. Thus, these teachers were able to map student scores to real-world skills development. In addition, they do an excellent job of showing how to deconstruct a nebulous construct into measurable elements. I was impressed with their thinking and their work. I was underwhelmed however by their resulting rubric. In short, they could use a good editor and they need to rethink the real purpose(s) and stakeholders for their rubric.

The audience for this article is clearly science teachers. It is written in a narrative style designed to appeal to practitioners.

Rubric Examples: Science Process Knowledge Rubric - Classroom teachers might have difficulty using this rubric for scoring their students’ work.


This translated article offered some useful elements that were not common to other articles in this bibliography. The study offered evidence that holistic rubrics were not very useful to low-performing students. They also found a significant amount of grader interpretation occurs when using holistic rubrics. More importantly, they identified several reasons for graders’ departures from the holistic rubric, including time, failure to concentrate, writer expression (I assume this means the expressiveness of the writing), attractiveness of the topic, differences between student
submissions, the sheer number of submissions, and situations where students of vastly different ability levels were all scored at the same time and by the same graders.

The list the authors provide here seems to be similar to anecdotal evidence I have collected in my experiences with analytic rubrics. So I support using this list to guide any rubric use.

The audience for this article is meant to be faculty members. However, the writing is slightly above the practitioner level. The work is further complicated by the translation effect. The rubric itself is a bit hard to follow as it jumps around in the sorts of criteria used and how those criteria are described. But there are some bright spots. Some of the approaches the authors used to describing quality offer alternatives to common methods seen in English-based rubrics published in the United States.

Rubric Examples: Multiple sections of an extended writing rubric. Note however that cultural differences yield different sorts of rows. The faculty greatly value some formatting issues (e.g., equal space above and below the text.


This article describes an excellent step-by-step approach for building rubrics. Like other authors, they point to the usefulness of rubrics as a learning tool. The authors describe using published standards from the start, justify the specific uses for their rubric, and then provide a good description of the steps they took in building their rubric. While methodical, the process has some bright spots. They realized that their criteria were too holistic and revised their rubric after the first use. They also made good use of pilot data for making revision decisions.

The authors suggest that multiple pilot tests would be quite useful. I agree. My experience has been that making a good rubric great only happens over time and over multiple uses. In addition, it really helps if new graders are employed to see the rubric with new eyes. The authors conclude that rubrics take a long time to develop, but that they are worth it for judging students’ ability to critique professional literature. I agree here as well.

The example rubric they provide has some interesting characteristics. Similar to the Renaissance Group’s Work Sample Portfolio rubric, these authors mix a checklist approach for some go, no-go sorts of measures followed by qualitative and quantitative criteria. This approach might be a good way to satisfy teachers who heretofore had only used checklist rubrics for things like format and design elements in their rubrics (e.g., used an introduction, presented subjects and methods sections, etc.). I’ve found that faculty members need a bridge to take them the next step into measuring qualities of the work. So rather than scrap the faculty member’s previous work, the format shown in this article could be used to augment the existing rubric.

The audience for this article is supposed to be other pharmacy school educators. However, the principles and reasoning in this article are quite similar to what many other authors offer to other audiences. I believe any reader interested in building rubrics might glean useful information from this article. What I like about this article is the practical approach – they had a very specific
problem that they needed to solve – and did. Unfortunately, their resulting rubric was not the greatest.

Rubric Examples: Students’ ability to critique professional literature. This rubric uses an interesting mixed format with a checklist followed by some scales.


This is a very short article that contains the results of a self-report survey of adult learners who used rubrics that had been handed out when their assignments were made. These adult students reported that they liked using the rubrics because the rubrics clarified expectations and helped them develop their assignments. These results are useful because they are consistent with other authors’ findings (e.g., Andrade, 2008). This article is in the review because it focuses on adult learners (rather than K-12 students) and because it provides support for the belief that giving students rubrics helps them succeed. This article provides some evidence that adults see some similar benefits to using rubrics during the learning process.

The audience for this article could be anyone interested in rubrics.

Rubric Examples: None.


This article provides a very different look at rubrics. Here the focus is on using them for evaluating textbooks based on the level of inquiry found within each of them.

The audience for this document is university science teachers. The authors do a fair job of outlining a very different development method and use for rubrics. What I like about this article is how they were able to make sense of an ill-structured problem and make use of rubrics as a tool for bringing structure and order to the situation.

Rubric Examples: A very basic quantitative rubric for scoring inquiry levels in textbooks. (Not for measuring inquiry in student submissions.) The poor rubric in this article would be useful as a “bad example” when teaching others how to build qualitative rubrics. However, for the purposes intended, the rubric is adequate.


I will begin this annotation by admitting that I am clearly biased toward the use of rubrics. Therefore, an emotional argument against rubric use did not sit well with me on first read. But as
I reflected on this essay I realized that it could serve as a cautionary message for those already well versed in rubrics. The authors provide a description of their emotional reaction, albeit based on partial knowledge of rubrics and rubric use. Their essay serves to show how negatively faculty can react to bad rubrics and bad rubric use.

The essay was written after a disconcerting conversation with a child about a homework assignment—a classic example of forming an opinion based on a sample of one. The authors’ comments are based on a bad rubric and on bad grading practices. For example, the rubric in question seemed to be very tightly defined with fairly trivial elements. The authors believed that the passing score set a very low bar for achievement. The authors suggested the teacher involved communicated the concept to students that the only way to pass was to follow the rubric to the letter. In essence, this entire situation was the antithesis of a good rubric and good rubric use. Other authors included within this bibliography make it clear that rubrics should contain criteria that matter, leave room for creativity, set high expectations, and be used in ways that encourage students to build better projects. Obviously, the child they mention didn’t experience any of this.

The audience for this essay is all educators. But I urge readers to first learn about the characteristics of rubrics and what they can do when used correctly before reading this essay. Look at the empirical evidence of the benefits of rubric use and examples of good rubrics. Also, as the reader tackles this essay and its emotional language, note how the authors tend to over-generalize from their sample of one to all rubrics and rubric uses. For example, they argue that the situation is going to affect “our entire society” (p. 199). They argue that teachers must “purposely set the bar low” (p. 199). Astute readers will come away from this article with a better understanding of why it is so important to build better rubrics and to use them in ways that help students learn and succeed.

Rubric Examples: None


This website contains a good solid set of steps for developing a new rubric. The steps are consistent with good practice. They provide a ten step process in straightforward language.

1. “With your colleagues, make a preliminary decision on the dimensions of the performance or product to be assessed.
2. Look at some actual examples of student work to see if you have omitted any important dimensions.
3. Refine and consolidate your list of dimensions as needed.
4. Write a definition of each of the dimensions.
5. Develop a continuum (scale) for describing the range of products/performances on each of the dimensions.
6. Alternatively, instead of a set of rating scales, you may choose to develop a holistic scale or a checklist on which you will record the presence or absence of the attributes of a quality product/performance.
7. Evaluate your rubric using the criteria discussed in Part 1.
8. Pilot test your rubric or checklist on actual samples of student work.
9. Revise the rubric and try it out again.
10. Share the rubric with your students and their parents.” (Multiple Web Pages)

This is a website designed for teachers in Chicago but the principles would work in many rubric development situations. What I like about this website is the simple step-by-step approach the authors use and the simplicity of the instructions. This is a good site for a beginner.

Rubric Examples: None (Very Surprising)

**Clauser, B. E. (2000). Recurrent issues and recent advances in scoring performance assessments.**

The author does a fair job of providing another way of looking at developing rubrics. His article is really about the broader issue of scoring performance assessments, but the concepts apply to rubric development. He organized his article around four questions:

1) What aspects are being scored?
2) What standards determine the score?
3) How are standards developed?
4) How are standards applied?

The author is clearly writing to measurement experts rather than to teachers. A novice in rubric development may not get much out of this article.

Rubric Examples: None.


The authors focus on how researchers might use the tool to help them extract meaning from survey data. The authors explain how they developed and used their rubrics. The steps they describe are adequate but other authors have better descriptions of how to build a rubric.

The issues discussed help us see how a very different purpose can completely change the nature of the rubric. In this case, the authors built very task-specific rubrics that would only be useful to them. Also, their description of what was difficult cautions readers about the time and effort needed to outline criteria and to write scoring descriptors for each criterion.

The audience for his article is not clear to me. It seems to be useful to other survey researchers.

Rubric Examples: An excerpt from a task specific rubric that might only be useful for the authors’ specific research effort.

This article is one of the better documents I reviewed. The authors do not delve into the relative merits of rubrics nor do they itemize methods for constructing them. Rather, they present important findings for those of us charged with measuring life skills like writing and critical thinking.

The authors begin by describing the difficulties of measuring the two nebulous and interconnected constructs of critical thinking and writing. They point out that a major problem we face is the lack of agreement about what either construct means. But then they turn to the data.

It seems that there is a major misconception that needs to be addressed. They found that high scores in writing were negatively correlated to scores in critical thinking. The reason, they speculate, is that the two constructs demand different conditions for proper measurement. Typical writing rubrics tend to measure standard writing abilities based standardized for collecting writing samples. However critical thinking demands a different sort of writing where students are free to be more creative and open. Thus, when students receive prompts aimed at measuring writing, the chances of their exhibiting any critical thinking within their submission is extremely low. Conversely, if prompts are aimed at getting the student to utilize critical thinking, their writing will not resemble what many consider to be “good writing.”

The audience for this article is higher education educators and academic leaders. I suspect however that this was also aimed at accreditation evaluators. Regardless, this is one of the more useful articles I have reviewed because it presents some important points not found elsewhere.

Rubric Examples: Not a rubric per se, rather a framework for building a critical thinking rubric. See the Facione (1994) article for a better look at critical thinking rubrics.


This is a simple and well done study that shows how faculty members conceptualized and built a rubric to measure accepted aspects of good flow diagrams. Their research methodology is sound. I will focus instead on the rubric development here.

The authors began their work by researching the literature to determine what constituted a high-quality flow diagram. They went beyond physical attributes to learn that good diagrams can show a student’s ability to transfer (what they call transforming knowledge, p. 48) their learning to their activities. As they continued their literature review three types of characteristics emerged for measuring high-quality flow diagrams including:

1. “Structure (completeness, appropriate sequencing, presence of subroutines where relevant)
2. Applicability (relevance of information, evidence of deep processing in translating information from manual to diagram (transfer), usability of flow diagram for guiding another student’s work without using the manual as well.

3. Feature of the Diagram (indication of quantities, correctness of representations of apparatus and/or processes, conception of how apparatus works, attention to detail such as annotation of the diagram with drops or motion indicators to denote shaking, use of arrows and numbers to fully describe the flow and action steps)” (pp. 48-49).

The faculty were able to achieve 80% inter-rater agreement when using their excellent rubrics. They found they were able see their students’ understanding levels and were able to provide excellent feedback.

The rubric they built was excellent. In addition, I found their explication of the rubric via two examples an enlightening section of the article. Their use of the actual diagrams and their interpretations brought the rubrics to life.

The audience for this excellent article is any faculty member interested in building a rubric for diagrams.

Rubric Examples: Flow Diagram Rubric – This rubric is adequate, but the diagrams that follow that help define the rubric are an excellent and highly recommended addition to any graphic rubric.


This is a fairly long descriptive article that explains how librarians at Washington State University implemented a scoring rubric for information literacy skills. What makes this article useful is their inclusion of a detailed rubric that is aligned to national information literacy standards.

The audience for this article is anyone building information literacy rubrics.

Rubric Examples: A high quality and detailed information literacy scoring rubric. I especially liked the way the authors defined their constructs at the tip of the page before presenting the rubric elements.


This article is useful in that it covered verbal communications whereas most authors discuss written communications. The authors adapted a rubric from one used by the National Communication Association to evaluate student performance in general public-speaking courses.

The authors introduce useful points that seem to be written for the benefit of accreditation team readers. For example, in the body, and then again in the discussion section, they point to the need
The authors highlight their use of external standards to ensure that “good” on their rubrics fit professional level work. This is excellent practice. They highlight the double-edged sword related to the number of scoring points across the rubric table. Fewer points (they used a three-point scale) made grading easier and improved inter-rater reliability – but this approach reduced the precision of feedback for instruction. Finally, they also brought up the need for significant grader training.

The audience for this paper is unclear. While it seems to have been written for higher education practitioners, it has a significant political undertone. Still, I found the article useful because they make a strong case for issues related to better rubric use.

Rubric Examples: None.


I won’t review these documents in this bibliography. I’ve placed these here to remind the reader that many excellent sources of rubrics and descriptions of the constructs they measure are available in the literature and online. The “Delphi Study” as the first report is nicknamed, is one of the best descriptions of a nebulous construct I’ve found.

Rubric Examples: A holistic critical thinking rubric.


This very short article presents an attempt to build a character development rubric. I reviewed this article with high but unfounded hopes. The author presents a poorly constructed rubric that contains many of the sorts of errors other authors caution against. In fact, I had to read the whole rubric multiple times to try to figure out how it might be used. This article remains in the bibliography because it represents an excellent bad example.

The audience for this article seems to be other elementary school teachers.

Rubric Examples: Character Development rubric. While the online format is ugly, the way the author made the constructs observable was quite good.

This article made me laugh out loud. I included this article in this bibliography to show that emotional essays (see Ghapman & Inman, 2009) can also come down in favor of rubrics and rubric use. The authors of this article are strong proponents of rubrics and rubric use. However, they make wild claims about the virtues of rubrics and overlay these comments with socially conscious claims about social science education. For example, they suggest that getting students involved in using rubrics will allow them to “live the social studies content” (p. 154). Teachers can benefit as well in that they can “align their curriculum and instruction confidently and competently” (p. 154). Thus, their students will feel “safe and wanted” because of democratic principles and social justice found in the classroom (p. 155).

The authors provide what they call “types of assessment” (p. 156). However this list is really a set of learning activity types (e.g., debate, blog, poem, map, poster, etc.). The rubric they offer is one of the worst I’ve seen in the literature.

The audience for this article is unclear. I believe it was written for other educators. However I included it in this bibliography only to provide a non-example showing how faculty with partial knowledge can embrace rubrics in their classrooms for odd reasons.

Rubric Examples: One of the worst rubrics I’ve encountered in the literature. It uses trivial criteria that are based on counting objects in a student submission.


When I first read this article I placed it on the “not useful” pile. However, on reflection, there is one gem buried in this otherwise lackluster piece. The author makes the point that involving those who will use the rubric in developing it should lead to better and more reliable use of the resulting rubric. She brings up “ownership” and the possibility of having graders gain better insights about students. I have seen similar outcomes when I involved graders in rubric development. They began talking about “our rubric” instead of “the rubric.” Graders also benefited in that they deeply engaged with the constructs being measured, which I believe helped them use the rubric more reliably. Note that I found similar suggestions in other articles. None were based on recorded events as was the case in this article.

The audience for this article is high-level practitioners who develop rubrics for scoring in large-scale implementations. However the reader may benefit from the one concept mentioned above.

Rubric Examples: None.


The authors add to the discourse on changes in the views of assessment. The authors begin with a description of how assessment has been shown to be the true or “hidden” curriculum (p. 3). Students will put effort into what they know will be assessed, often to the exclusion of all else.
Their review of the literature found evidence that multiple choice exams and surface essays tend to guide students into surface learning and cramming at the end of a semester. Both of these behaviors are good strategies for getting grades, but poor learning strategies. Instead, when assessments are designed for engagement (e.g., performance assessments with good rubrics) then students tend to go into deep learning mode. The primary value of good assessment however is that it provides meaningful feedback—which has been shown to have a stronger effect on learning than any other tool in the teachers repertoire (see Black and William, 1998). The key elements of successful assessment and feedback are listed.

- Sufficient number of assessment tasks drives sufficient student study time and engagement.
- Distribute assessment over the course of the semester.
- Align the assessment task to the outcomes directly.
- Include meaningful, detailed and appropriate feedback at every opportunity.
- Feedback should support further learning of the assessed task.
- Feedback should be communicate clearly with the students.
- Design ways to ensure that the students can (and do) use feedback. (pp 12-25).

Rubric Examples: None


This article is not directly related to rubrics per se. Those interested in rubric development methods won’t find help here. Instead, this article is about measurement principles that can guide rubric development. The author does a good job of outlining the need for meaningful cognitive models if we want to build better ways to teach and test thinking. The key ingredient is to find ways to deconstruct thinking into component parts. The author outlines how difficulty (complexity, abstractness, etc.) plays a critical role in how the construct is manifested in life, and therefore in leaning and assessment.

The audience for this document is well beyond the average practitioner. It seems to be aimed at those with a good background in educational and measurement theory.

Rubric Examples: None.

**Griffin, M. (2009). What is a rubric? Assessment Update, 21(6), 4; 13.**

I have read hundreds of ponderous and stuffy definitions of rubrics. This was not one of those. The author of this piece didn’t forget “rule six” – (don’t take yourself too darn seriously)! The result is a very brief and readable article that explains what rubrics are.

Some of the more memorable parts of this article included these observations:

- “A rubric is more like a cake than a rock. It contains whatever we put in it, not what nature designed.”
- “A rubric is a series of choices. Unfortunately, like a phone booth, it can be stuffed with only so many bodies.”
• “The signers of the Declaration of Independence could not have spent more time haggling over words than most committees charged with creating a rubric. Yet the discussions that haven’t killed it have made it stronger.”
• “Like a living document [rubrics] must change with the times.”
• A rubric is “what we promise to teach.” (all of the above quotes – pp. 4; 13)

The audience for this enjoyable article is anyone interested in learning about rubrics. I highly recommend this article.

Rubric Examples: None.


This is a very detailed report of a fairly good research study. The authors’ literature review makes the point that literature prior to 2003 consisted mainly of narratives and speculative pieces rather than empirical studies of rubrics. My own literature review revealed the same trend. Their rubric is designed for scoring an oral presentation which is also worth noting because most authors cover written communications.

What makes this article especially useful to me is the finding that with proper rubric development and use it is possible to see congruent ratings between faculty and their students. I must admit my preconception that students would score quite differently. However the empirical results suggest that if the authors’ methods are used, student grading is a viable option for use in higher education. The results clearly support the idea that using rubrics for education could extend from formative self-review to operational peer review scoring.

The catch however is that a few conditions must be in place to expect congruent scoring. First, students needed to be involved with rubric development. Thus, they “owned” the rubric – meaning they felt it was theirs, and they understood it. Second, the students were “additional” or “second” scorers because the instructor marked every oral presentation and also asked students for their confidential and candid scores (p. 1513). Thus, students experienced some pressure to conform to the rubric to ensure they were consistent with the teacher’s scores. Note also that the author admitted finding four “bogus” scores (p. 1525) where students gave themselves maximum points and others significantly lower scores. Thus, any attempt to use this technique would need safeguards of some sort. If I were to use this method, I would reject the highest and lowest scores for each presentation before computing my averages.

The audience for this article seems to be college teachers who are well versed in rubrics and various methods for scoring. The point the authors make about peer review (and the empirical backing for their comments) make this a useful article for anyone learning about rubric use.
Rubric Examples: A poorly constructed oral presentation rubric. Note that “collaboration” would be very difficult to score because graders might not see these behaviors and other rows in the table are broad and subjective.


Like other authors in higher education over the past several years, accreditation (accountability) pressures drove Halonen, et al, to build rubrics and assessments to cover their accreditation standards. The resulting rubrics cover broad skills that reflect those standards:

- Description
- Conceptualization
- Problem Solving
- Ethical Reasoning
- Scientific Values
- Communication
- Collaboration
- Self Assessment

I recommend these rubrics to the reader because the developers attempted to create scales that measured abilities across the four-year degree program.

The audience for this article is meant to be science teachers. However, I believe it would be a good article for any teacher who has some background in curriculum development and rubrics. The article is quite good for a couple of reasons. First, it contains a very detailed set of rubrics for the previously listed skills. Second, it details very specific benefits and shortcomings of using rubrics.

Rubric Examples: Good rubrics for several skills – see bullet list above. I was impressed with how the authors built real-world ability/expertise scales for their rubrics.


The authors of this article do an excellent job of picking apart some common assumptions about portfolio assessment. This document is included within this bibliography because nearly everything the authors discuss actually pertains to the use of scoring rubrics for evaluating portfolios. Their study offers evidence that pure holistic scoring was not very successful for their faculty. They learned that they needed to define specific criteria for scoring to overcome difficulties such as lack of standardization of what constituted a successful student submission, and what criteria should be used for making grading decisions.
The authors found that nearly all their graders jumped to a holistic score soon after reading the only the introductory paragraphs of a student’s submission. Thus, the assumption that having more data to work from (by reading the rest of the student submission) could improve the reliability of grading was not supported in practice. In fact, graders who did go on to read additional material in student submissions found that it served to confuse their decision making.

The authors also learned that requiring student submissions from various genres did not enhance the grading process. More submissions from different genres served to confuse the graders. Note that the authors were quick to point out that there are curricular benefits to requiring multiple genres that may outweigh the difficulties in grading.

The authors further encountered a surprising result. They learned from their graders that the quality of a student’s writing was fairly consistent across genres and across multiple submissions. Unfortunately, it was not clear from the article whether these findings were a product of inappropriate grading behaviors—such as jumping to a score for a submission before reading the entire document—or due to underlying systemic factors such as genre or multiple submissions.

The findings in this article are of great importance because they sharply contrast with commonly held beliefs about holistic rubric scoring. More and different submissions did not improve grading. The study also pointed to the need to move away from holistic scoring toward analytic methods with more explicit response criteria. Building analytic rubrics enhanced their graders “standardization” and allowed them to improve their grading overall. The authors speculate that using more explicit criteria could be a tool for forcing graders to attend to each student submission. They also suggest that using slightly different criteria for each genre may help graders attend to each of those submissions as well.

These findings clearly point to the need to reduce the cognitive load for graders. If rubric developers don’t reduce the load, graders may use inappropriate strategies to reduce the load themselves. This conclusion is consistent with other discussions of analytic scoring that suggest more explicit criteria reduce the cognitive load on graders.

Another important finding in this study is that scoring the process students use to complete their submissions (as opposed to simply scoring the resulting products) is feasible. Further, they found that identifying any improvements in a student’s writing over time is also possible, but requires explicit rules and prompts to guide the nature of the student submissions.

Late in the article, the authors fell off the track. They suggested that portfolio assessment was an excellent tool that “builds consensus in assessment and instruction” (p. 186). However, a careful read of the authors’ reasoning reveals that rubrics and portfolios are not the vehicle in this case. Rather, collaborative discussions among faculty members about the criteria for scoring the portfolios builds consensus. Building rubrics can help faculty bring issues out in the open, grapple with what they value, foster deeper communication, act as a vehicle for training, and—in some cases—help faculty come to some measure of consensus about what constitutes exemplary student performance.
The audience for this article is supposed to be English composition teachers. However, this is a valuable article to anyone who wants to learn about how to build and use rubrics.

Rubric Examples: None.


The authors offer a very brief look at building a rubric for their counseling program. Like other authors, they bring up important aspects such as careful definition of criteria, using the rubrics during instruction, and cautioning readers about how time intensive building rubrics can be. They offer a step-by-step approach but provide very little detail.

The audience for this article is clearly others in their profession; especially those who have little to no knowledge about assessment and rubrics. As such, this might be a good beginner’s article.

Rubric Examples: Excerpts of their Clinical Supervision “Rubric.”


This article is mostly a political piece aimed at communicating with accreditation organization committee members, school administrators, and politicians. The article seems to be an affirmation of outcomes assessment standards and rules. As such, I didn’t find this article very useful for understanding rubrics, but it does highlight the heavy influence of accreditation and other external standards on current practice.

The real value of this article to anyone learning about rubrics is in the many rubric examples that the authors provide.

The audience for this article could be educational practitioners.


The authors give us a very good look into their thinking as they developed a rubric. Important concepts included such things as how criteria ought to be identified, how to brainstorm and define the criteria, how to refine the criteria, and then how to use the criteria in scoring and grading. The authors suggest that “in the absence of criteria, assessment tasks remain just that, tasks or instructional activities” (p. 2).
The authors are quite specific throughout. They went into greater detail on how rubric developers ought to consider the needs of their stakeholders – more so than most of the other authors I’ve read. For example, they have separate sections discussing how rubrics should connect with instructional developers, parents, and students. They found that criteria can support instructional planning by helping teachers define instructional goals and that rubrics are of value to students because they can engage the students in deeper discussion of the criteria for an assignment. The authors also found that rubrics helped parents support student learning from home.

Then the authors outline a simple set of steps for beginning rubric developers to follow when specifying criteria. However, this list is far too simple to yield good results.

- “List multiple learning outcomes.
- Divide each outcome into performance levels.
- Describe traits/characteristics for each level.
- Provide a numerical scale for rating” (p. 5).

Later in the article, the authors do present some useful questions rubric developers might ask themselves as they specify criteria.

1. What are the attributes of the performance that matter to me?
2. What qualities or features will help me know whether students have produced an excellent submission?
3. How does completing this task relate to my goals for students?
4. What samples do I have of excellent work?
5. What made these samples excellent?
6. What dimensions and/or criteria appear in my state curriculum or other standards? (Paraphrased from page 8).

The authors argue that good criteria should communicate to all stakeholders, help the educational process, support reliable grading, and focus on valid issues. They also argue (as I do) that good criteria emerge over time. Even after the rubric is “done,” pilot testing will inevitably result in changes.

I especially liked the author’s description of high quality criteria.

- Linked to important student learning outcomes (all the important ones).
- Are sensitive to the sorts of decisions you will make with the rubric.
- Are credible, fair, and unbiased (based on accepted, real-world attributes of quality over which students and teachers have control)
- Are feasible (limit the number and complexity)
- Yield meaningful and clearly understood scores.
- Offer concrete references that are clear to all stakeholders. (Adapted from pp. 12-14).

The audience for this article is anyone who actually has to build criteria for a rubric. I found this article very useful.

Rubric Examples: None.

The authors of this excellent article discuss how they use rubrics as an integral part of their instructional approach for teaching preservice teacher candidates about potential bias in assessment. The approach is based on concepts found elsewhere in this bibliography.

Students learn that rubrics are useful, motivating, and powerful tools but that it is very easy to insert their biases into their rubrics. Thus, validity can suffer if rubric developers and users are not vigilant in their work. In fact, if rubrics are built using the sorts of rules presented in this bibliography they can help teachers avoid bias. A few points worth noting include:

- All assessments are biased because they reflect the values of the designer.
- These biases extend to the sorts of standards or other resources we might use to develop our rubrics.
- Carefully constructed rubrics can reduce bias by making our sometimes subconscious biases visible and allowing the developer to explicitly deal with them as they select criteria and develop the descriptors.
- Standardizing and carefully following the grading process can reduce bias that normally creeps into any human endeavor.
- Using proper grading practices can reveal our internal biases about “good” or “bad” students as we compare their work to fixed criteria. (The authors point out how surprised some candidates are when they see how well their “poor” students actually perform and how poorly some of their “good” students perform.)
- Properly designed rubrics have the potential of helping students learn and teachers teach by revealing the students’ strengths and weaknesses.
- Education can sometimes be improved simply because students have a better understanding of their grades, and the transparency leads to more positive attitudes about fairness.

The audience for this excellent article is supposed to be preservice education faculty members. However, I believe anyone who has a basic understanding of pedagogical practices and rubrics would benefit from this article.

Rubric Examples: An adequate lab rubric. Be careful to remove any mixed aspects when differentiating between a score of 3 and 4.


I included this online resource because it provides a basic look at rubrics for beginners. Also, it provides useful criteria for good rubrics. Other articles in this reference list provide much more detail. If a reader wants a good starting point for learning about rubrics, this might be it.

The audience for this website is beginning rubric developers.
Rubric Examples: A rubric template with suggestions.


This was a very difficult article to read. It was clearly written to well-informed instructional designers. The author used high-level jargon (e.g., *amplitive skills*) and it assumed a significant background in educational theory. I remember some of what they discussed from previous reviews of learning theory literature.

What I did find useful was how the author broke out broad constructs like problem solving into various sorts of elements (e.g., concepts, rules, mental models, applying arguments, and inference). The point he makes is that broad constructs must be defined before they can be used.

The audience for this article is an advanced practitioner who is comfortable with learning theory.

Rubric Examples: None.


This article is one of the most useful I’ve found because it reviews 75 other research studies that each offer evidence for the various benefits of using rubrics.

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. (Proper rubric development is a key means for achieving validity.)
3. Rubrics seem to have the potential of promoting learning and/or improve constructivist 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” (p. 130).

The article is written above the practitioner level. It is aimed at instructional designers and those who are experienced with rubrics and education. I found this to be a useful article because it added evidence to some key points (listed above) about rubric development.

Rubric Examples: None.


If the reader is looking for a lightning rod paper, this is it. Kohn’s essay is an emotional argument against the use of rubrics and grades. The author “expresses doubts” (abstract) about rubrics in a column entitled “Speaking My Mind” found in the *English Journal*. 
The author makes several broad generalizations, offers opinions, and provides negative comments about education and rubrics. For example, he feels that the correct purpose of authentic assessment “must be the elimination of grades” (p. 12). Thus, rubrics are exactly the wrong things to use. He then suggests that building consistent criteria is not possible. He says that rubrics are designed to change teachers into nothing more than grading machines and schools into “test preparation factories” (p. 13). He feels that standardizing in any way compromises the quality of teaching and the quality of learning.

A counter point to this essay can be found in the Spandel (2006) essay reviewed later in this bibliography. Also see Mabry (1999) for a differing view on these issues.

The audience is other English teachers.

Rubric Examples: None.


The authors offer a comparison between two rubrics designed for measuring critical thinking from Facione (discussed earlier in this bibliography) and Newman (not reviewed in this bibliography). The authors feel that both rubrics measure a similar conception of critical thinking. This is an important point because critical thinking is a particularly disputed construct.

The authors found that the Facione rubric is fairly general and holistic whereas the Newman rubric is extremely detailed and specific. Also, the Facione rubric uses a four point scale whereas the Newman rubric uses positive and negative indicators that look very much like traditional text content analysis tabulation. The study found that the Newman rubric was too detailed, too complex in actual use, and took an inordinate amount of time for graders to use. On the other hand, the Facione rubric turned out to be too general and holistic to provide meaningful feedback.

These findings are important to us in that rubric developers need to provide enough detail to address the measurement purposes for the rubric but need to avoid getting too mired in unnecessary detail. A good rubric has to balance between overly analytic and detailed criteria and overly holistic criteria.

The audience for this article is educational practitioners.

Rubric Examples: A very difficult to follow (and use) Newman Team Indicators of Crucial (Critical) Thinking.

When I first saw this article, I thought it might be another piece that simply argued against measurement and standardization (see Kohn, 2006). Instead, the author does a good job of outlining what can go wrong with rubrics. He successfully argues that over-standardization is just as bad as not having defined criteria because it fosters formulaic writing. He points out how using good standards as a starting point is fine – as long as developers take care to operationalize the standards that they value in good writing. He points out that careful consideration about what students should know and be able to do should be considered a positive rather than a negative. The author does a very good job of showing the relationship between validity and reliability. If rubrics are too standardized and too detailed, they may lack validity. The job then is to balance the needs through careful selection and definition of the criteria at just the right level of detail.

The author is clearly writing to teachers. The article provides not only cautionary comments about bad rubrics, but offers solutions to build good ones. The downside of this piece is the author’s lack of citation of the research in defense of many of her points (although in fairness, much of the literature does not emerge until after publication of this article).

Rubric Examples: None.


The main focus of this article is on how to define and measure “interdisciplinary writing” in higher education. But the article clearly describes how the authors took the nebulous construct and defined a “framework” for understanding and measuring it.

The authors begin by making the argument I’ve made elsewhere in this bibliography, namely, that people like Kohn (2006) and other authors who argue that rubrics are promoting shallow learning are basing their positions on bad rubrics and bad rubric uses. Available empirical evidence suggests that good rubrics and proper rubric use can lead to better learning.

The authors begin their construct work by reviewing the literature on interdisciplinary writing (IW) to find a common definition and outline potential aspects of the construct. They determined that for their purposes IW would be the “capacity to integrate knowledge and thinking in two or more disciplines or established areas of expertise to produce a cognitive advancement – such as explaining a phenomenon, solving a problem, or creating a product – in ways that would have been impossible or unlikely through single disciplinary means” (p. 337). (Quoting Boix, Mansilla, Miller, & Gardner, 2000 – not reviewed in this bibliography.)

The authors go on to build a three-part framework for assessing IW that looks at disciplinary expertise, integration of two or more disciplines, and making informed selections of material from each discipline to address their needs. Then they build definitions of four levels of “understanding” ranging from *Naïve* to *Novice* to *Apprentice* to *Master*. Finally, they build a series of “questions” that stand as descriptions of each criterion. I list these issues to try to show
the reader how deeply the authors delve into the nature of the construct of interest. They don’t settle for surface qualities. Instead, they dig deep to find the underlying structures that best represent quality.

The authors go on to tell us that their specific criteria are their values and that others might use the same framework but find different criteria based on differing values. This is a critical point for new rubric developers to understand. Even a simple construct can be seen differently by various teachers. If an institution expects to build a rubric that will be useful across a campus, then it should be founded on a firm framework that guides the nature of the rubric scores yet leaves room for disciplinary (and frankly individual teacher) differences. We use this type of development at my campus. We use a template that outlines a framework of eight general criteria categories. However the descriptors are left to individual departments to define. For example, quality writing for nursing students has some unique attributes compared to quality writing for biology students. Even within a criterion (a row on our rubric table) like “Credible Evidence” we see slightly different definitions of what “good” looks like for each type of writing. Yet, because both groups are looking at “Credible Evidence,” the results can be compared.

The audience for this article is anyone interested in building an interdisciplinary writing rubric or any cross-course/curriculum rubric.

Rubric Examples: None.


This article is helpful because it presents a different use for rubrics. Rubric development can guide the definition of nebulous constructs for just about any purpose. The authors use rubric development as a vehicle for curriculum development. The authors begin with the nebulous construct called “a rigorous curriculum” and break that down into its component parts. The result is a four-level rubric that provides a detailed description of poor to great rigor in curriculum, instruction, and assessments. Once they had the rubric in place they move on to create detailed questions faculty should ask themselves. Then they move forward to build concrete examples of good curriculum.

The audience for this article is anyone interested in curriculum review. Those who are experienced with rubric development will benefit from seeing this very different use for rubrics.

Rubric Examples: An interesting (and almost holistic) Academic Rigor rubric.


This is a straightforward study of inter-rater reliability using a middle school mathematics scoring rubric. The authors found significant error in scoring using their rubric. They speculate that the nature of the tasks and halo effect both played a part in the differences in scoring.
The reason I included this article in this bibliography is to point out that the authors did not delve into the effects caused by the nature of the rubric itself nor the training they offered to graders. They attempted to use a general rubric across many tasks. The rubric was fairly complex and a little hard to follow. It focused almost all of its descriptors at the top scoring level while giving short shrift to lower scoring levels. Also, their scoring scheme seemed a bit weak to me. They were not very firm about what constituted given outcomes. Thus, I believe that improving the rubric and scoring scheme might yield better results.

The audience for this article is intended to be math teachers. However, methodological problems suggest this article should be used by readers with a background in rubrics and scoring.

Rubric Example: An adequate middle school mathematics scoring rubric.


This is another good solid introduction to rubrics. The author presents a very readable step-by-step process for building rubrics followed by a couple of fair examples. Mertler is a strong proponent of analytic rubrics. He correctly highlights the value to students in detailed feedback that offers strengths and weaknesses on specific attributes of the submission. He does a better than average job of explaining how to convert rubric scores to grades using logic—versus averaging or percentage points. The only real weakness is the brevity of the step-by-step section of the article.

The audience for this document seems to be elementary school teachers. The process described and the principles of a good basic rubric seem to be useful to others as well.

Rubric Examples: A basic albeit mediocre rubric for scoring research papers. There are other examples as well, but none are especially good.


This is a book on “authentic” assessment. The author offers some fairly good chapters (I reviewed chapters 5-7) dealing with rubrics, rubric development, and methods for using rubrics.

In chapter five the author makes the point that rubric development can be a stepping stone for better definition of new assessment tasks that are aligned to the learning outcomes. She argues that this linkage—instruction to assessment—is an essential step in becoming a better teacher. In chapter six the author argues (as I do) that rubric development is often nothing more than documenting what we as teachers “already do in our heads” (p. 56). Later in the chapter she provides several example rubrics to help the reader see many formats. In chapter seven the author highlights the differences between measuring product versus process and then she focuses on how to use rubrics to help students self-assess as part of their learning process.

What I especially liked about this book was the quality of the *many* example rubrics showing many different types designed for very different purposes. I also quite liked the last chapter that
showed how to adapt scoring rubrics into student self-scoring rubrics designed to improve learning.

The audience seems to be consistent with the title—elementary teachers, but could just as easily be used by anyone building rubrics. It certainly reads like a textbook.

Rubric Examples: Templates of Holistic and Analytic Rubrics; A poorly constructed “Group Oral Presentation” rubric, and a mediocre Research Paper Performance Task Scoring Rubric. Many of the rubrics in this text are not very good. For example, the “Rubric for Visual Presentation” (Figure 6.1) offers comparative language in the scoring and requires written feedback. The worst example in the book is found on page 71. The “Listening Rubric” contains several measures that a grader would not be able to observe – thus rendering them nearly impossible to measure. There is one highpoint. The “Rubric for Teamwork Behavior” on page 72 contains some excellent concrete descriptors.


The author presents some basic information about rubrics and then offers several examples of the rubrics her institution uses. In her next section she outlines the Green River Community College Campus-wide Outcomes. Interestingly, they are written to be very specific so that building criteria (rows in her rubrics) is quite easy and alignment is easily verified. The balance of the document contains instructions to faculty at her institution—including a survey for judging the quality of their own rubrics, a listing of active verbs, and more rubric examples. The only major weakness I see in this article is that the author’s use of verbs is rather confining to building better descriptors of each scoring level.

The audience for this document is faculty at Green River Community College but it might be an easy read for any teacher. Also, the document is aimed squarely at supporting accreditation in that the author uses examples of outcomes and rubrics for their campus-wide outcomes.

Rubric Examples: Many. Some examples utilize an odd scoring method and mix aspects across the rows. The “Oral Communications” rubric is among the best on this website. The site contains a metarubric of sorts – really a survey about rubric qualities. In addition it contains examples of rubrics for: Written Communication (both holistic and analytic versions); General Communications; Mathematical Concepts and Procedures; Group Discussion; Reasoning Ability; Oral Communications; Mathematics Extended Response Item Rubric; “Single Competency Rubrics” including Organization; and Sentence Fluency.


Few articles discuss rubrics designed for scoring creative or graphic submissions.

The graphic submission rubric the authors generated was based on their values about the environment so the criteria reflect the sorts of content they expected to see on students’ drawings.
rather than the quality of those drawings. This outcome supports my belief that all assessments reflect the values and biases of the assessment developer. For example, in this case, the authors value receiving pictures that contain illustrations of humans interacting with the environment. Another set of authors might value drawings that focus more on cycles in the biosphere or conceptual models of pollution.

I included this article because it illustrates an important concept for rubric developers. When we build rubrics we need to carefully consider what matters in the students’ submissions. I have seen how novice rubric developers sometimes jump to easy things to measure such as format, color, spelling, use of headings, etc. when the real value of the submission is to show whether the students understand a complex issue or concept. Thus, it is important to consider the real purpose for the submission and the value to the student’s education before settling on rubric criteria.

The audience for this article is anyone who is interested in building a rubric for measuring the content of a graphic submission.

Rubric Examples: An excellent Draw and Environment Rubric (measures the content of a hand-drawn picture).


The author presents multiple types of rubrics and mentions when each might be an appropriate tool. I included this article because the author touches on descriptive versus comparative judgments in rubrics. The concept here has real value to anyone learning to build better rubrics. The author suggests writing rubrics that describe what the student’s submission will look like instead of writing descriptors that list inferential judgments about the relative quality of the work. For example, we would look for writing that was “free of noticeable errors” rather than writing with “minor errors.”

The audience for this article is slightly above educational practitioner level. The reader may need some background in rubrics to fully benefit from this article.

Rubric Examples: Holistic Writing Rubric.


In this very brief article the authors describe how they went about building a writing rubric in their chemical engineering department. They also discuss how to tie accreditation (and other) standards to rubrics. They present a useful description of thinking deeply about a set of standards to yield meaningful criteria. I thought that their discussion about the choice of a scoring scale that supports meaning and reliability was quite good. Their resulting rubric is usable.

The audience is other teachers.
Rubric Examples: A poorly constructed Engineering Student Mathematics Application rubric. The trouble with this rubric is that some of the aspects across the rows would be extremely difficult to discern from a student submission.


This set of presentation slides is a weak attempt at teaching people how to build good rubrics. What is strong and worthy of inclusion in this bibliography is the author’s description of the thinking one must go through to develop criteria of “goodness” for a given purpose. She uses “good dogs” as her example which greatly improves understanding of the complex process of selecting the correct criteria for inclusion in each row of a rubric.

The audience for this slide presentation is any faculty member interested in developing a rubric from scratch.

Rubric Examples: Sadly, none.


This work is an article (and associated presentation slides) that describe issues related to rubric development. What makes this piece worthy of inclusion in this bibliography is the very different sort of rubric aimed at information literacy. I caution the reader however, to look at this only in general terms to see the sorts of criteria the author selected to address specific standards. I don’t recommend attempting to use the author’s development process; moreover, her statistical analyses are problematic.

With the noted caveats (above) the audience for this presentation is anyone interested in building a rubric based on student activities related to information literacy.

Rubric Examples: Some mediocre rubrics dealing with information literacy standards. The problem with the examples is that the authors tried to stick too closely to the language in their standards.


The authors present an excellent example of using theoretical literature as a way to break down a complex nebulous construct like critical thinking. Compare this method with other approaches to selecting rubric criteria such as simple brainstorming or just listing what we value. I found the authors’ analysis of the literature on critical thinking useful. They decided to limit their
definition of critical thinking to “argumentation” and outlined three outcome measures for this—including use of evidence, use of research, and the ability to use other perspectives.

While their description of their construct analysis is good and their methods for limiting their final criteria are solid, their resulting rubric is poor. So readers are cautioned to focus only on the first ten pages of this article.

The audience for this article is unclear. The title suggests faculty in a preservice teacher education program. But the article seems to be written for someone with significant background in critical thinking, rubric development, and preservice teaching requirements.

Rubric Example: Yes, but I had difficulty making sense of their poorly constructed rubric.


This article discusses a new approach to using item response theory (IRT) parameters for discovering troublesome performance tasks and rubrics. The technical details are not germane to this bibliography. However, the author’s discussion is quite enlightening in that they outline several issues they found in tasks and rubrics worthy of mention here. I will translate their discussion into rules for rubric developers. Suffice it to say that the IRT work the authors presented offer support to these issues.

1. Carefully specify the criteria in scoring rubrics.
2. Align the criteria to the essential concepts underlying the content area of interest. (This has to do with discussions of Construct Irrelevant Variance (CIV) in scores. That is, if the criteria measure things that are not central to the quality of the work, then the resulting scores will contain variance that is not relevant to the task at hand.
3. Use an appropriate number of scoring levels to capture the meaningful differences in performance. I would add that you should use the minimum number possible that still capture the nuances in quality of student performances. The reason is that too many score levels leads to trivial difference between levels that are hard for graders to differentiate. On the other hand, too few scoring levels leaves some performances in between the lines with no way to score them.
4. We know that tasks should give students a way to show their reasoning and understanding. Rubrics must provide a means for scoring the various ways students might show that understanding. For example, in math, some students will use equations while others will use the written word to show their understanding of a concept. Both are valid and so the rubric should allow for these two types of communication.
5. Base criteria and scoring on good evidence found in the literature and standards.
6. Make the criteria and scoring developmentally appropriate. For example, expecting younger students to use high-level mathematical notation – when they are still barely capable of explaining math concepts in words is inappropriate.
7. When a general rubric is used, verify that it is capable of capturing all the performances for each task it is to score. For example, if an assignment is different than all others, and it has different expectations, then you may need to adapt the general rubric to fit.
8. When writing descriptors to explain each scoring level for a criterion, ensure that the descriptor for a “3” that is supposed to mean “meets expectations” actually defines that level of quality. This ensures consistency and interpretability of the scores.
9. Ensure that graders see all levels of quality for every criterion during their training.
10. Continuously improve the rubric by adding more and better detail to the descriptors (or graders notes) to further explain the differences in scoring levels and to better capture all student performances.

The audience for this article is measurement experts. Readers who don’t have a firm background in IRT should only scan pp. 218-221.

Rubric Examples: None.


Rubrics dealing with musical performance are not common in the literature. The author of this article does a good job of explaining how he went about conceptualizing a fairly good rubric for measuring musical performances in his department.

The author guided the project using “assessment in a learning culture” (p. 98) which comes from Lori Shepard’s (2000) article (not reviewed in this bibliography). The author decided to build a rubric and grading system that:
1. Was settled [placed] in the middle of the teaching and learning process (not at the end).
2. Provided feedback to the students.
3. Offered significant scaffolding for learning.
4. Offered new (novel and un-encountered) tasks to encourage transfer.
5. Used firm and published (transparent) criteria for measuring excellence.

The authors review the literature and found the sorts of things we see in other articles in this bibliography. They decided to build analytic rubrics because holistic rubrics tend to lead rubric users (graders or scorers) to adapt the resulting vague rubrics to their own definitions of success (see p. 99). Also the author greatly values formative feedback so he wanted to build a rubric that would offer rich and constructive help to students.

The author collaborated with other faculty to develop detailed common rubrics that they later tested in a pilot study. They learned that students have a positive reaction to rubric use that likely supports learning because students clearly understand what they need to change to improve their performances.
The author concludes that his efforts were successful, met the six characteristics of a good grading system (see above) and moved away from the “secret nature” of musical performance (p. 104).

The audience for this excellent article is meant to be other music faculty. However the principles and descriptions found here would be useful to anyone interested in building educative rubrics.

Rubric Examples: A solid musical performance rubric.


The authors discuss their university's implementation of new types of assessment and reflect on the process and their findings. The authors discuss how critical thinking has been accepted in standards, but that clear definitions of the construct and good exams are not in place. They used this perspective as they reevaluated the situation and selected criteria for their rubric. I especially appreciated how the authors compared their work to the accepted literature.

The article seems to be written for business faculty. However, I suspect it was also written for accreditation committee members. The author’s even use the accreditation/accountability phrase “closing the loop” and go on to describe what their faculty are doing to close that loop.

Rubric Example: Critical thinking rubric. While the format is difficult to follow, and the resulting rubric is mediocre as built, the reader may find some useful elements that could be reused in other critical thinking rubrics.


The author seems to wander among several related topics in this article. However she has some material that may be of value to certain readers. She begins with a description of how to select performance tasks. While not rubric development, I have seen faculty confused by which assignments they ought to use to measure certain constructs. Pearlman offers a way to think through better assignment selection. This alone makes this article a useful document. Next she offers an adequate tour of holistic versus analytic and also general versus task specific rubrics. The following section is an excellent set of steps faculty should use for selecting a preexisting rubric for their own use.

*Note: This article seems to be a near duplicate of a chapter she wrote for a book published in 2002. I reviewed the chapter version because it was easier to read.*

The audience for this article is unclear. Given the wide range of topics, I believe it is not targeted at any one group. Those who are just getting into measures that go beyond multiple choice tests may benefit from one part of the article, while only those who manage large-scale grading might benefit from the final section of the article.
Rubric Examples: None


This was a very different sort of article than most I reviewed. The authors discuss the development of a rubric designed for program-level assessment. They invoke Peter Ewell (a known expert in accreditation and assessment) in the first line of their article. Disappointingly, the authors talked a lot about what should happen – rather than getting very specific about what they did. Also, their resulting rubric showed a lack of understanding of good rubrics. For example, their scoring levels were comparative and lacked detail. Thus, inference levels would be quite high. I kept this mediocre article in this bibliography simply to show the reader how a rubric could be used for things other than scoring student submissions, and to use their substandard rubric as a non-example.

This article is very clearly aimed at accreditation committee members.

Rubric Examples: A Poorly constructed Project Assessment in Management of Business Information, and very poorly constructed Systems Analysis & Design rubric. For example, in Appendix 2, the rubric contains comparative language leading to high levels of grader inference. In addition, the rubric contains rows where the only difference between a three and a four is an added aspect that was not a part of the rest of the row.


This useful article describes how the authors developed a scoring rubric for case-based narratives. The authors offer a strong description of general versus task-specific rubrics in their review of the literature. Then they describe how they surveyed the literature to identify potential criteria for their rubric. Based on pilot data, the authors shifted their task-specific statements to more “generic” (p. 84) descriptors to try to better fit their assignments and scoring needs. They also described their version of scaling for scoring the results. While I disagree with their scaling methods, I believe this article will help readers see how simple rubric scoring can be modified to fit theoretical and practical purposes. Unfortunately, the resulting rubric is not very good. It is strong on counting things and weak on qualities of the students’ work.

The audience is anyone with a good foundation in rubric development methods and uses. A basic understanding of scaling and composite scores would be useful but not required.

Rubric Examples: A lackluster task specific and generic rubric.

The author does an excellent job of outlining elements of good and bad rubrics. This brief article is an easy read for anyone learning to build better rubrics. The author’s thesis really boils down to making the criteria too specific and/or detailed versus too general. His target, however, is classroom rubrics so the reader might need to generalize his comments to fit a rubric designed for larger purposes. For other information on the concept of general versus detailed criteria descriptors, see Landis and others (2007) and Mabry (1999).

Popham is a very down-to-earth guy who has told me that, above all, he values simple clear communication in his writing (personal communication, 2008). Popham certainly does not disappoint, as just about any educator or lay person might understand this article.

Rubric Examples: None.


This is a fairly lengthy report of a study that investigates the role of benchmark writing samples in assessing writing and examines the consequences of differential benchmark selection with a common writing rubric. If the reader is interested in the methodology and details, please do read this work. But I can save the reader some time here. The most useful message for those building rubrics is that it is very easy to bias the results through selection of benchmark (A.K.A. exemplar) papers to define scoring levels.

What the authors do not cover is how to resolve this potential source of error when using rubrics. The key is to develop low inference rubrics with very strong descriptor statements across each row to define scores. Do not leave the full definition of scores to the exemplar papers. It is far too easy to bias the results if the rubric developer failed to explicitly describe what each score meant in the first place.

The audience for this paper is anyone learning to build rubrics. However, the paper might serve to confuse beginners who may not see the connection between strong descriptors across each scoring row and the less explicit means of defining scores known as exemplar sampling.

Rubric Examples: None.


This is a technical report of a study that has one very important message for rubric developers. We must take great care and put much thought into how we define each scoring level represented in a rubric. A method that is used for holistic grading is called anchor papers. These are full
student submissions that represent various score points on the rubric. The authors found that how
graders interpret scoring levels and the resulting scores can vary based on which papers are
selected as anchors. They also found that when using a rubric across multiple ability levels (such
as a rubric used for novice-through-expert students) the problem can increase substantially.

The concepts presented in this article have value to analytic rubric development as well. How we
go about defining each scoring level—and for whom—can greatly affect the resulting scores. If
exemplar submissions are used, they should be very carefully selected to represent specific rows
of the table to improve precision. Finally, rubric developers may want to consider how their
ability scales work across ability levels. For example, is the rubric that is aimed at determining if
a freshman is a decent writer really a good measure of senior writing?

The audience for this very technical article is any measurement expert.

Rubric Examples: None


In this excellent article, the author outlines six rules for developing criteria for student
submissions. But more importantly, the author outlines his specific reasoning behind each rule
and offers examples to illustrate his points.

1. Significance – criteria important to successful performance in the real world including
   content, cognitive, metacognitive, and dispositional components.
2. Fidelity – criteria are nearly identical to criteria for successful real-world performance
   and use standards that match how successful performance appears in the field.
3. Generalizability – criteria that are just right for addressing the purpose of the scores. The
   appropriate level of specificity is determined by how tightly the criteria relate to the given
   task versus all similar tasks. For example “the student has a valid conclusion” versus “the
   student summarizes valid traits.”
4. Developmental Appropriateness – criteria should cover the expected levels of
   performance for the students being tested.
5. Accessibility – the language used in the criteria should communicate clearly to students.
6. Utility – how well the criteria profile meaningful strengths and weaknesses that can be
   affected by instruction. (pp. 320-331.)

The rules in this article are similar to the concepts presented in the late Samuel Messick’s
seminal work on test validity. I won’t review that chapter here as it is written for very advanced
measurement experts. Suffice it to say that Quellmalz’s ideas are quite consistent with accepted
best practices in the measurement world.

Although the reading level seems to be slightly above classroom practitioners, I think the content
makes this article a must read.

Rubric Examples: None.

This paper is actually more of a proposal for a future study. Its strength is in the author’s review of the existing literature. He examines the impact of rubrics on student learning and seeks to explore the usefulness of rubrics from students' perspectives. He is also interested in using rubrics for examining student attitudes and inclinations as well as their utility for making curricular decisions.

The author explains that rubrics are many things such as: a scoring schema, a means of communicating expectations, a way to provide feedback to students, and a source for grades. He lumps the conceptualization and operationalization of criteria descriptors into one step – which I have found does not work very well. He argues that criteria descriptors could take the form of “levels of mastery,” performance descriptors,” or “target indicators” (p. 5). He points to the literature to remind rubric users that rubrics designed to provide feedback are also useful for grading and for looking at course or program level outcomes. He goes on to argue that there is sufficient literature to suggest that rubrics can overcome the two most prevalent student complaints—lack of fairness and lack of useful feedback (p. 5). He also finds literature to support the idea that multiple faculty can agree on a rubric for use across sections of a course or even across multiple courses – though he does admit that we need to leave room for minor differences for each assignment. Finally the author explains the concepts of metacognition and self-regulated learning and how these are guided by a student’s motivation, self efficacy, and other contextual issues (p. 8). The author closes by lamenting the “dearth of empirical work” (p. 9) supporting many of the issues listed above.

The best part of the document is the author’s conceptual model for studying the benefits of rubrics.

Rubrics can be considered “good” when they are:
- Reliable because they have explicit criteria
- Valid because they are linked to student learning outcomes (I would add – standards and/or external criterion-referenced ability scales)
- Effective in providing transparent, consistent, detailed feedback.

Rubrics are useful for student self-assessment when they are used for:
- Self-monitoring ongoing efforts to complete an assignment
- Accurate self-assessment of performance
- Guiding revision

Rubrics have affective benefits if they are perceived by students as tools for:
- Clarifying the purpose of the assignment
- Explaining the nature of the grading
- Providing a fair measure of their performance

Rubrics can support better learning if their use leads to:
- Enhanced engagement
• Heightened interest
• Stronger effort
• Higher confidence
• Better grades

I sent a query to the author to see if the study was eventually funded. I didn’t receive a response.

Rubric Examples: None.


The authors present a critical review of the rubric literature. However, they limited their work to twenty articles that contained empirical evidence. Even so, their findings are similar to what I found as I developed this bibliography that covered many more articles and books.

A caveat: The reader is cautioned to see these pieces of evidence as supportive, but inconclusive. That is, rubrics, and rubric use in education looks promising. Most of the data to date has been quite positive. However relatively little solid research has been published to date.

Of benefit to students, empirical evidence suggests that rubrics may:

- Define success
- Clarify assignments
- Help students focus on the important aspects of quality in their work
- Offer transparency and fairness to the grading process
- Facilitate self-regulation in the learning process
- Improve the quality of student submissions

Of benefit to teachers, empirical evidence suggests that rubrics may:

- Promote learning and achievement – especially deeper learning
- Define and measure difficult constructs such as critical thinking and reflection
- Provide much needed feedback on student work
- Be used to scaffold difficult concepts
- Be used to foster deeper class discussion
- Improve the consistency, accuracy, and efficiency of grading
- Quickly identify group strengths and weaknesses to guide remediation
- Provide data useful for making curricular and pedagogical changes to improve learning
- Be used to gather program-level assessment data.

All of these benefits assume that the rubrics are good ones and that they are used appropriately. Readers should review articles that contain meta-rubrics to see what constitutes “good” rubric design. Of special note, rubric developers need to fully define only the criteria that matter most and then they should use the resulting rubric as an educative tool rather than just as a scoring and grading tool.
The audience for this article is anyone researching rubric development and use. However, a good background in rubric development and use would be very helpful for understanding the complex issues the authors present here. This is the best review article I uncovered in my review of rubric literature.

Rubric Examples: An underwhelming annotated bibliography rubric that has some confusing descriptors that seem to wander as you read across each row in the rubric.


This is one of the very best articles for readers interested in learning how to build better rubrics. The article goes into great detail about rubrics and exactly how to build them. I especially liked the author’s discussion about the thinking behind each step and what a good outcome looked like. The article also outlines very specific general and technical rules for good rubrics.

The author informs us that the best rubrics:
1. “Are sufficiently generic to relate to general goals beyond an individual performance task but specific enough to enable useful and sound inferences on the task
2. Discriminate among performances validly, not arbitrarily—by the central features of performance, not by the easiest to see, count, or score.
3. Do not combine independent criteria in one rubric.
4. Are based on analysis of many work samples, and based on the widest possible range of work samples—including valid exemplars.
5. Rely on descriptive language—what quality, or its absence, looks like—as opposed to relying heavily on mere comparatives or value language (e.g. “not as thorough as,” or “excellent product”) to make the discrimination.
6. Provide useful and apt discrimination to enable sufficiently fine judgments—but not using so many points on the scale as to threaten reliability (typically involving, therefore, 6-12 points on a scale).
7. Use descriptors that are sufficiently rich to enable student performers to verify their score, accurately self-assess, and self correct.
8. Highlight the judging of the “impact” of performance—the effect, given the purpose—as opposed to over-rewarding merely the processes, the formats, or the content used; and/or the good faith effort made.” (pp. 1-2)

I especially liked how the authors delineated different types of criteria in a list that went beyond the typical product versus process and the quantitative versus qualitative dichotomies. They made a good case for criteria that measured:

- Impact (effective performance)
- Craftsmanship (quality)
- Adequacy of Process & Behaviors
- Aptness or Validity of the Content
- Degree of Mastery or Sophistication (adapted from headings on p. 5).
The author argues (as I do) that beginning rubric developers often focus too much on content, format and conventions while missing impact and methods (p. 5).

The authors then proceed to present an excellent and detailed guide for developing criteria and descriptors that any novice rubric developer would benefit from following.

The audience is any educator interested in building a rubric. This article was very readable and outlined in very specific ways exactly how to build a great rubric. This is perhaps the very best article for explaining exactly how to develop rubric criteria and descriptors.

Rubric Examples: Several. Note: the reader should be aware that many of the examples are non-examples that the authors use to illustrate rubric errors. Do not use the examples in this text as “good” examples! Narrative Scoring Guide, Advanced Placement U.S. History Essay, 5th Grade Inquiry: Oral Presentation; Open-ended Math Problems, Argumentative Composition, High School History Senior Essay, Virginia Literacy Passport, Socratic Seminar Performance Assessment, Mathematics (Problem Solving, Reasoning, Effectiveness of Work, Accuracy of Written Work, Quality of Presentation, Clarity); New York Writing Handbook, 5th Grade Science Project, Spanish Proficiencies, Upper Arlington Development Reading Scale, 1st Grade Reading, UK Writing Rubric, NAEP Mathematics Proficiency.


This is not an article. It is a part of a “Portfolio Toolkit” I found on the Internet. These pages present an excellent “metarubric” that readers could use for evaluating the quality of a rubric. I disagreed with the author’s decision to build a holistic metarubric. Although the rubric on this webpage is good, it seems to me that judging specific qualities of a newly constructed rubric should be done with an analytic metarubric to provide stronger feedback for improving the draft rubric.

The audience for this document could be any educator.

Rubric Examples: Holistic Metarubric.


This book offers a detailed description of a multi-year (and ongoing) effort to build 15 rubrics that cover most of the key enduring skills that educators value. The rubrics themselves were conceived as broad institution-level descriptions of performances (see p. 2). The rubrics have an interesting metric. The lowest end of each scale [a score of “1”] represents entry level performance for freshmen and the top end reflects the most complex performances rather than a senior level performance (p. 3). The AAC&U cautions us not to read too much into the scales.
The four-point scales don’t correspond to years in college nor do the entry-level scores represent college readiness (page 3). The resulting AAC&U VALUE rubrics “reflect faculty shared expectations for essential learning across the nation regardless of type of institution, mission, size, or location” (p. 21).

In chapter three the authors summarize the multi-year process of developing their rubrics. I found it interesting that they admitted they had false starts and had to learn the best ways to do this. Also in chapter four, the authors begin by saying that few faculty have any experience with building or using rubrics (p. 15). This is interesting because I have seen exactly the same difficulty here at my campus. The authors explain how they had to create their own ways to build their rubrics because of a relative lack of literature on the topic. I found this consistent with what I found the first time I developed this literature bibliography. Until only recently, few authors offered books or articles about analytic rubric development and even less on how to develop scoring rubrics for broad enduring skills.

In chapter five the authors present a protocol for using the rubrics in their own teaching. I won’t repeat it here but I will mention some highlights that I didn’t find in other authors’ works. First, maintaining a glossary of terms used in the rubric would help grading. Our graders use this approach frequently. Second, the protocol stresses how faculty can remain true to the intent of the rubric.

The text is very readable and straightforward—even if chapter two shows a level of crafting by a writing instructor who is fond of sentiments like “a rubric is a series of choices…unfortunately, like a phone booth, it can be stuffed with only so many bodies” (p. 9). Note that this is not a problem; I actually found her expressive (perhaps even lyrical) writing about a technical subject refreshing. The concepts presented in this book are consistent with other authors comments on the value of rubrics to the educational process (see most of the other books and articles in this bibliography).

The audience for this book is just about any teacher who is interested in using rubrics in their classrooms or across their institutions.

Rubric Examples: Many excellent – albeit rather holistic – rubrics (e.g., Civic knowledge & engagement (local & global); creative thinking; critical thinking; ethical reasoning & action; foundations & skills for lifelong learning; information literacy; inquiry & analysis; integrative learning; intercultural knowledge & competence; oral communication; problem solving; quantitative literacy; teamwork; and written communication).


The authors explain how they used educational theory and past research findings to define their rubric. The article is a good example of breaking nebulous constructs down into component parts and selecting what aspects of each are valued enough to measure. Although the authors offer a follow-up article on validating their rubric, this first article covers the important material for those interested in building rubrics for measuring nebulous constructs.
This document is written for advanced instructional designers. I think the article would not be useful to most teachers.

Rubric Examples: Rubric for Assessing Interactive Qualities in Distance Education Courses. The format of this rubric would make it difficult to use.


This is a useful article that presents a teacher’s use of a rubric to resolve an educational problem. Her students were not offering meaningful conclusions in their lab reports so she determined what she wanted to see in her students’ writing (e.g., a focus on the purpose of the lab study, written in technical language, etc.) and then developed a rubric that provides good examples of unidimensional scales across the rows. Her results were encouraging as her students began submitting conclusions that better matched the intent of that section of their lab reports.

I liked this article for a couple of reasons. First, it was quite practical and straightforward. The author had an educational problem to solve, found that better scoring would help, and built a good rubric for accomplishing the task. Second, she gives actual examples of student work and the respective rubric scores.

The audience is clearly any teacher or faculty member. I highly recommend this article.

Rubric Examples: Lab Conclusions Rubric. This is one of the best rubrics I’ve reviewed. It offers concrete and meaningful description on each of the rubric rows.


This is a brief and simple article in which the authors suggest that clear, accessible rubrics give students practice with planning, revising, and editing while helping them become self-regulated writers through self-assessment and peer assessment. In other words, rubrics are good for scaffolding learning.

The authors offer a very positive opinion about using rubrics in writing instruction that is a near polar opposite of Kohn’s position (2006) (discussed earlier in this bibliography).

The audience for this article is just about any teacher or faculty member.

Rubric Examples: A student version of a 4th grade rubric for a mock interview of an historical person. The authors present a nice use of first-person descriptors for helping young students engage with the rubric.

I admit my biases. I adore rubrics because I’ve seen them do so much good for students and faculty so reading articles that naysay rubrics gets my back up. Although this author attacks analytic rubrics, he presents a logical review of issues that helped me see some chinks in the pro-rubric armor. While I didn’t agree with certain points in the article, and I didn’t care for his approach of looking for which type of rubric is better (holistic or analytic), overall his work brings out important issues that ought to be considered when building and using rubrics.

The author begins with a clear thesis: analytic rubrics “can, and in many cases, do, fail to meet the conditions for sound assessments of complex student works” (p. 159). His description of the nature of rubrics as subjective measures of divergent student works is spot-on. His separation of quality of a whole student work versus individual qualities (criteria) is quite useful. His depiction of holistic and analytic rubrics is generally good.

The author assumes that each type of rubric is used exactly as planned. But my experience suggests that the relative benefits of each type of rubric ought to be compared against actual—instead of optimal—design and practice. For example, limitations in a faculty member’s ability to articulate and operationalize what is valued may hamstring analytic rubrics—thus leaving important criteria unstated, and therefore unused. On the other hand, holistic scoring can be hamstrung because valued criteria are purposely left in the grader’s head and gut rather than placed on the page, which can lead to other sorts of errors.

I did find two specific points where I disagreed with the author.

1. I soundly disagree with the author’s contention that analytic criteria (each row in a rubric table) should be mutually exclusive. I believe the opposite must be true to have any hope of the sum of the parts being similar to a measure of the whole. Analytic criteria should be mutually supporting, such that only the major foci should change between rows. For example, the concept of building a written document that has good content development is certainly attached at the hip with the concept of making good claims or providing good credible evidence. Further, from a statistical viewpoint, the criteria ought to co-vary.

2. I also disagree with the author’s negative comments about the inclusion of a holistic or overall measure within an analytic rubric. Perhaps my disagreement is due to the fact that I use the holistic row differently than the author’s described use. When students have made a few errors on specific criteria, but the work as a whole stands together, then the holistic measure can (and should) be used to bump the decision to passing. Note—I don’t use my holistic measures to bump scores below the passing point.

Although the article is an attack on my favorite rubric type, it is of real value because it outlines many important aspects of rubric quality and best practices that any good rubric developer ought to be concerned about.

1. Consider the value of feedback in the educative process and build your rubrics to address this use. Then consider how to reuse the same data for scoring and grading.
2. Carefully consider the overall purpose for using the rubric before considering the criteria or scoring-level descriptors.
3. Use real-world standards and human judgment to create a better list of the most important criteria for scoring. To this I would add – collaborate among peers to create a stronger criteria list and try to “standardize” what quality means across sections of a course, courses, or even across the entire campus.

4. Dig deeply into the criteria so that the rubric includes hard to measure constructs rather than settling for what is easy to count.

5. Consider the size of the “chunks” on each criterion row. If you go too wide, then you are actually building a holistic rubric. If you go too narrow, then reductionist principles apply and you may end up with trivial measures. This is a critical and often misunderstood concept that deserves special attention.

6. Don’t expect rows to be mutually exclusive but do be sure the focal point of each row differs sufficiently to allow separate measurement.

7. Reduce the inference in the rubric by creating better definitions of each criterion. I prefer to use descriptors of what success would look like instead of using verbs that create a relative scale (e.g., minimal, some, most, etc.).

8. Train graders to follow the rubric. For example, don’t allow redistribution of scores to fit a holistic judgment. But try to give graders some flexibility for dealing with odd submissions.

9. Pilot test the rubric to find missing criteria and to improve the descriptors to better capture all student performances.

10. Do not expect interval level grading scales. Ordinal is as far as human judgment can take you.

The audience for this article is well beyond the practitioner level. Readers might need significant background knowledge in both holistic and analytic rubrics to be able to glean useful information from this work.

Rubric Examples: None.


This article is a straightforward research report that explored “the effects of teacher knowledge of rubrics on the achievement of high school students” (p. 151). Findings suggest that knowledge of rubrics may be beneficial. However, the findings were equivocal because of methodological issues and because only one of the content areas (biology) showed a positive outcome.

The audience for this article is above the educational practitioner level. Researchers and measurement experts might get more out of the study. I included this article in this bibliography because I can use their substandard rubrics as non-examples when teaching others how to build better rubrics.

Rubric Examples: Poor examples of rubrics. Do not use these rubrics as exemplars.

The author offers a description of rubric building for her music class. She takes the reader through some of the thinking she used and steps she performed. This article is important for a couple of reasons. First, the author presents a useful argument for task-specific rubric design (as compared to general rubrics designed for scoring multiple assignments). The concept here is that the rubric should be made relevant to the assignment and to the students who will create submissions. While absolutely true, the same is also true of general rubrics. I should note that it is certainly easier to align task-specific rubrics to these two concepts of good rubric design. Second, the author does a great job of explaining the difficult concept of construct irrelevant variance (CIV) in scores. She explains that the scoring must “represent student learning adequately in relation to the underlying skills and competencies required for completion of the task” (p. 10). She goes on to say that it is important to omit “aspects of the performance that are extraneous to the underlying skills and abilities of the students” (p. 10). This is the classic definition of CIV. Third, the author advocates first defining skills and competencies and then developing an “instructionally relevant task” (p. 11). Interestingly, the author invokes the late Samuel Messick’s seminal work on validity. It is clear to me that she understands measurement theory quite well.

The audience is supposed to be any music teacher. However, the author’s description of guarding against construct irrelevant variance would be useful to anyone already familiar with the basic steps for rubric development.

Rubric Examples: A fifth Grade Music Class Rubric for Assessing Rhythmic Compositions. This is a poor rubric but has value because it offers good ideas for criteria in a music-related rubric.


This article immediately caught my imagination because as a parent I have often struggled with helping my children with their homework. The author offers anecdotal evidence in support of her belief that using rubrics communicates the intent of assignments to parents. Specifically, parents are able to use the information found in rubrics to better understand how to help their children.

I included this article because it offers a different look at instructional uses for rubrics. In this case, the rubrics are aimed at increasing parent understanding of the assignments in order to increase the likelihood of their getting engaged in the learning process with their children.

The audience for this article is any teacher. See also the Whittaker (2001) article for more about the use of rubrics for communicating with parents.

Rubric Examples: 5th Grade Research Project.
This is an interesting article that discusses some important aspects of good rubric design. The authors cover building scales that make sense, how to specify anchors, and analytic versus holistic rubrics. Then the authors discuss something rarely mentioned elsewhere – the use of content versus construct aspects within scoring rubrics. This is an important distinction that new rubric developers would be well served to understand. They could use rubrics to judge the quality of writing and/or the content of that writing such as understanding of a complex historical issue.

The authors describe not only how they constructed their rubric, but also do an excellent job of describing how they created criteria and scales before having access to any exemplar papers.

The audience for this article is unclear. I believe much of what the authors discuss would be useful to most rubric developers. However, they do explore details about validity and constructs that may require some experience with these concepts. Unfortunately, their rubric has some major flaws.

Rubric Examples: A substandard portfolio assessment rubric. For example, on the “Accuracy” row, the authors slipped in the word “concise” to help delineate a score of four from a three. Conciseness has little to do with accuracy.

This article is in this bibliography because every analytic rubric development project breaks constructs down into measureable elements. Rubric developers need to be very careful to not reduce the construct down so far that they destroy the meaning of the original work. While reliability is a good goal, taking it too far can damage validity and may lead to trivial measures.

The audience for this article is above the educational practitioner level. Readers should be familiar with the measurement concepts of reliability and validity and should understand the basics of performance assessment.

Rubric Examples: None

This is a very simple study that really has only one message – students prefer to have feedback rather than a single holistic score. Although obvious, it is useful to have empirical backing for this concept.

The audience is just about any teacher.

Rubric Examples: Excerpts from scoring guides.


This is a simple report of a research study that provides support for the use of rubrics to improve learning. The authors found that shared scoring guides (A. K. A. *rubrics*) reduce writing apprehension, act as a substitute for teacher-student conferences, saves faculty time and effort over written feedback or conferences, urges students to participate in the writing process, and tends to reduce student complaints.

The audience for this article is just about any teacher. What makes this fairly simple study meaningful is that it provides empirical evidence to support the idea that rubrics can support learning to write.

Rubric Examples: None.


This article supporting rubrics appeared in the *English Journal* just a few months after the Kohn article (2006) article that lambasted rubrics and measurement in teaching writing. The author explains that is it not the rubrics and measurement that are the problem. Instead, it is what faculty place in the rubrics—what they value—that is the problem. When faculty members reduce writing to measures of format and syntax, then higher-level aspects like organization and impact can be lost. Therefore the author believes that “when rubrics are thoughtfully crafted and used with discretion and understanding, they can be among the most useful instructional tools that teachers have” (p. 19). She goes on to argue that the underlying issue is fairness. I would add that another underlying issue is the faculty member’s expertise with making what they value measureable. Measuring higher-level issues like cohesiveness or addressing the proper audience is certainly harder than measuring format and spelling.

The audience is any teacher. For a broader pro-con viewpoint, it would help if the reader perused both this article and the Kohn (2006) article.

Rubric Examples: None.
The authors provide a solid text on rubrics and rubric development. This is one of the better texts available to helping readers learn to create rubrics. Topic coverage (and length) is not as exhaustive as the Wiggins (1998) book but it contains a significantly more material than the websites and articles listed elsewhere in this bibliography.

In chapter one the authors discuss the purposes, types, and parts of rubrics. In chapter two they respond to the title “why use rubrics.” Rather than simply list why rubrics should be used (timely feedback, etc.) the authors provided real-world situations to illustrate why faculty might want to use rubrics. I found this list very useful.

- “You are getting carpal tunnel syndrome from writing the same comments on almost every student paper.
- It's 3 A.M. The stack of papers on your desk is fast approaching the ceiling. You're already three weeks behind in your grading, and it's clear that you won't be finishing it tonight either.
- Students often complain that they cannot read the notes you labored so long to produce.
- You have finally graded all your papers and worry that the last ones were graded slightly differently from the first ones.
- You want students to develop the ability to reflect on ill-structured problems but you aren't sure how to clearly communicate that to them.
- You give a carefully planned assignment that you never used before and to your surprise, it takes the whole class period to explain it to the students.
- You have worked very hard to explain the complex end-of-term paper; yet students are starting to regard you as an enemy out to trick them with incomprehensible assignments.
- You work with your colleagues and collaborate on designing the same assignments for program courses, yet you wonder if your grading scales are different.
- You are starting to wonder if they are right.” (p. 4-5).

Chapter three offers a general process for rubric development, but then more specific construction methods are offered in chapters four and five. Chapter six covers uses of rubrics and seven offers several very useful examples of rubrics designed for specific purposes. Finally, the authors provide many appendices to provide even more examples.

The audience for this text is anyone who wants to learn to develop rubrics.

Rubric Examples: Many examples. Some of the examples are holistic rather than analytic rubrics. The “Film Presentation” rubrics have a very interesting checklist-within-analytic formatting. The lab experiment rubric however is nearly holistic and should be split into several rows to make the rubric more usable. The graphic design rubric (Figure 7.7) is a bad example in that none of the scoring levels are defined. Appendix F contains a very bad example rubric because it offers the grader a confusing layout and little to differentiate between scoring levels.

The authors offer a straightforward description of a classroom need and their process for developing a scoring rubric to address their need. Their methodology is generally sound and very readable. They offer some good advice along the way.

The authors went an extra step that I didn’t see in other authors’ articles. They carefully defined each scoring level before building the rubric. For example, on a four point scale, level three represented first draft but acceptable work that didn’t need revision and a score of four represented excellent work. Level two, on the other hand, suggested that revision would be needed because only partial understanding was exhibited. Those receiving a score of two or one would require remediation before attempting a second draft.

The authors reminded readers to consider common misconceptions as they define their rubrics. They also point out that rubrics define a standard and that grading on the curve is no longer an option. They suggest that their methodology gives teachers a better understanding of what they expect of their students and gives them a way to communicate that to their students.

The authors present a method for student revisions that avoids cookbook copying. Students must turn in their first draft, the scored rubric, their new draft, and a description of what they did differently to address the shortcomings in the first draft. In an afterthought they mention that using the rubric improved grading and reduced grading time.

The discussion the authors provide is quite good because it outlines a solid process for defining what student should know and be able to do in mathematics.

The audience for this article is any group of teachers.

Rubric Examples: A very interesting graphic formatted rubric. The authors’ “Grading Rubric for Metamorphosis Paper” is interesting in that it uses a checklist within a nearly holistic rubric (see Figure 5.1). However their Appendix K presents a “diversity” rubric that would be very difficult to use.


This article covers exactly what the title implies. The authors successfully argue that inconsistency in rubrics is a common error that must be overcome to make the rubrics more useful and fair. They provide a very detailed description of what good/bad looks like and how to fix or avoid making errors. The do an excellent job of describing scales (rows on a table) that measure just one thing. They also provide a good list of potential attributes for good rubrics:

- Breadth
- Accuracy
- Relevance
• Clarity
• Importance
• Impact

Please note that the authors focus heavily on building parallel wording across each row rather than focusing on defining degrees of the attribute on each row. Also, they focus on relatively abstract attributes rather than pressing for concrete ones. In fairness, perhaps this is a place for novice rubric developers to begin? I highly recommend this article to anyone learning to write rubrics.

The audience for this article is just about any teacher.

Rubric Examples: Metarubric and many non-examples with suggested corrections. For example, Table 1 contains poor comparative descriptors. But, the format that shows the criteria and its definition right in the rubric table is an excellent method for communicating to stakeholders and reducing error in grading.


Rubrics can support purposes beyond student submission scoring. In this article the authors describe how they created a program-level rubric designed to guide curriculum development (e.g., courses, assignments, assessments).

The faculty collaborated to define core research skills based on their experience and upon discipline-based standards. Then they built scales for each skill to define Beginning, Novice, Competent, and Proficient performance. I especially liked how they started by characterizing the nature of each of these four levels (columns in a rubric table) before they attempted to build descriptors for each scoring level on each row. For example, a Novice level is characterized as research having “insufficient depth” or research that represents “incomplete knowledge of the area of interest” (p. 24).

The result was a comprehensive rubric that covered a developmental sequence (across the rows of the rubric table) for many key research skills (down the rows of the rubric table). The rubric forms a “unifying structure” (p. 26) for thinking about every aspect of the program. The authors use the same rubric for several purposes:

• Design courses, assignments, and exams.
• Score individual course assignments.
• Rate applicants before admittance to the program.
• At the end of each course, use the rubric to show progress on each skill (as compared to the scores from the previous semester).
• Determine the relative contribution of each course to given skill development.
• Foster student metacognition and reflection on their learning.
• Provide a fixed standard for progression through the program – thus documenting student and program success.
The audience for this article is anyone interested in building a rubric that contains a true developmental sequence that could be used for multiple program and student measures.

Rubric Examples: A clinical research mastery rubric.


The article discusses rubrics in inclusive classrooms. The majority of the article plows similar ground as many other articles in this bibliography. It describes the process for building rubrics and offers rules. In addition it makes some useful points about using rubrics to communicate expectations (benchmarks) for each element of an IEP and using rubric construction to help define the IEP elements. Like Shaw (2004) the authors suggest that rubrics can be used to communicate with parents.

The audience is supposed to be elementary school teachers. However, I feel that the principles would be useful to just about any teacher.

Rubric Examples: Elementary School Website Assignment. This is one of the worst rubrics I’ve reviewed. It contains mostly comparative (relative) scales and contains lots of counting.


This is an extremely useful book for teaching teachers how to build assessments. The most useful chapters are two and then five through seven.

Chapter two discusses “authentic assessment” (p. 21). The author explains how assessment should be based in actions very similar to those performed on the job or in life. He rightly points out the difference between an educational activity and an assessment activity aimed at collecting learning data.

Chapter five covers standards and criteria. The author explains that there are three types of standards including content, performance, and task-specific work elements. He goes on to list expectations for good assessment such as credibility, usefulness, fairness, rigorousness, and feasibility. Then the author goes into a detailed description of breaking a construct down into domains, tests, tasks, criteria, and finally, indicators (A. K. A. descriptors). The author lists types of criteria including:

- Impact of the performance
- Work quality and craftsmanship
- Adequacy of methods and behaviors
- Validity of content
- Degree of expertise (p. 130).

I like the author’s list of examples of each criteria type found on page 131.
Chapter six covers individual performance assessments. I found his description of the process for converting an educational activity into an assessment quite useful. While quite complete—perhaps even exhaustive—the template for defining a performance task was far more extensive than most faculty would ever use. Similarly, his checklist for development was also a bit lengthy for everyday use.

Chapter 7 finally gets to rubric development. This chapter is quite good. It covers the process for developing rubrics and important considerations as the process moves forward. I appreciated the author’s description of how to build rubrics. But more importantly, his comments about characteristics of them would be very helpful to anyone learning to build rubrics. For example, he does a good job of explaining how each row in a rubric should define a developmental sequence so that feedback would be useful to the learning process.

I found one minor issue to disagree with. He explains the selection and use of exemplar papers. However, those who are building a new rubric may not yet have access to exemplar papers. It is essential for those in this situation to understand how to build grader’s notes to further define scoring. The author’s final section in this chapter contains a list of the top characteristics of good rubrics.

What I liked about the book was that it provides very detailed step-by-step instructions and reasoning for each step in the process. I must admit however that I disagreed with some of his points of view and some of his specific methods.

Although written for classroom teachers, anyone interested in rubric development could benefit from this book.

Rubric Examples: Many excerpts used to illustrate points. I especially like his rubric on page 136 “Assessing for Contextual Realism and Authenticity.”


These two documents are descriptions of how to measure “better” or “critical” thinking. These documents appear in this bibliography because they contain specific instructions for using rubrics for measuring thinking.

The audience for these two documents is anyone interested in using rubrics to measure student thinking.

Rubric Examples: Better Thinking Rubric. The approach the authors took with this rubric is quite interesting. They built a progression of evidence rather than levels of quality or evidence. However, the rubric would be difficult to use as formatted.

This is a short and simple study report that has just one important finding. The modality for training rubric graders does not make much difference in outcomes. However, online training can be slightly faster than other modalities.

The audience for this article is researchers or anyone interested in building a large-scale grader training program.

Rubric Examples: None.