

The Official (and Unofficial) Rules for Norming Rubrics Successfully

Claire Holmes and Megan Oakleaf

Academic libraries provide value to their institutions in a variety of ways. Perhaps the most important, and most pervasive, is the contribution libraries make to student success. To demonstrate their impact on student success, academic librarians must assess the results of student interactions with the library. Among the most important student library interactions is information literacy instruction; therefore, it makes sense for librarians to examine students' information literacy learning. Whether information literacy instruction is course-integrated, offered through credit-bearing courses, or provided via optional workshops, assessment of student learning is vital. Thus, librarians who seek to connect libraries to student success must provide evidence of student learning by regularly and systematically assessing information literacy skills demonstrated through student academic work.

Because librarians recognize the need to conduct assessments that link libraries to student success, librarians “embrace the challenge of demonstrating the effectiveness of their instructional programs and partnerships” (Gilchrist & Oakleaf, 2012). Increasingly, librarians are learning assessment skills and forming partnerships with academic disciplinary faculty and assessment professionals. In addition, librarians are seeking tangible, efficient, and flexible assessment tools that are grounded in information literacy learning outcomes, but are able to accommodate multifaceted information literacy instruction programs; diverse library instructors, methods, and assignments; and various preferences of collaborating faculty and other stakeholders.

Rubrics are just such tools: they can be used as road maps for librarians, academic disciplinary faculty, and assessment professionals to examine students' work relative to information literacy learning outcomes. There is growing evidence that suggests that rubrics are, in fact, effective drivers of the complete information literacy instruction assessment cycle (ILIAC), because they offer a specific and systematic way to examine student learning outcomes and a method to examine tangible evidence of that student learning (Oakleaf, 2009a). However, rubrics do have a significant drawback; they are only as good as the raters

using them. To achieve consistent and reliable use of a rubric among numerous raters, and to create the best possible tool with which to examine student work, the rubric and the raters must go through a “norming” process. The norming process is usually lengthy and often complex; consequently, it benefits from the guidance of a facilitator. Librarians who know how to facilitate the norming process can approach rubric assessment of information literacy skills with confidence.

Imagine...

You and a group of your colleagues (perhaps a mix of librarians, academic disciplinary faculty, and assessment professionals) decide to assess student information literacy learning outcomes. As a group, you and your colleagues identify a set of student work samples that provide evidence of student information literacy skills and set about collecting those student work samples. Some member of the group (maybe you!) drafts a rubric to assess those student work samples. (For more about composing rubrics, please see the book entitled *Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback, and Promote Student Learning*, by Danielle Stevens and Antonia Levi.) You and your colleagues then begin to apply the rubric to the work samples, but quickly realize that, despite your best efforts to select appropriate work samples and compose a useful rubric, the group does not always agree on how to score individual work samples using the rubric. Because you know that inconsistent (unreliable) assessments cannot be accurate (valid) assessments, the group decides to address the disagreements through engaging in “norming” the rubric. You have been asked to facilitate the norming session, but you feel uncertain about how to proceed.

Fortunately, rubric norming typically follows a process (the “official rules”). Adhering to this process makes it more likely that you and your colleagues will achieve consensus when scoring student work samples with a rubric. In addition, through engaging in RAILS research at ten institutions nationwide, we have also learned some additional lessons about what to do, and what not to do, when norming a rubric (the “unofficial rules”). We offer these rules, both official and unofficial, to guide those who seek to facilitate the rubric norming process.

Unofficial Rule A: Someone has to be in charge.

Rubric norming requires a facilitator. Rubric norming facilitators need to understand at least four things: the rubric, the student work samples, the raters, and the rules of norming a rubric. Frequently, rubric norming facilitators are selected for this role because they have been involved in the composition or development of a rubric. In these cases, the facilitator is already familiar with the rubric itself. If this is not the case, then the facilitator should seek to learn about the rubric origins, including who wrote it, what their goals were, and what decisions they made when writing the rubric and the rationales for those decisions. If earlier versions of the rubric exist, then the facilitator should be familiar with those as well.

The facilitator must also be familiar with the student work samples to be assessed. Typically the best way to do this is to examine all of the work samples under consideration. In fact, we recommend that facilitators informally score all the work samples themselves, more than once, before attempting the norming process. While this process may be onerous, it ensures that the facilitator is fully aware of the range of work samples and the challenges they present, such as null, wrong, or “creative” student work.

A rubric norming facilitator who knows the raters s/he’ll be working with is at an advantage. Facilitators should learn all they can about their raters’ disciplinary affiliations, experiences with assessment, connections with the student work samples, and prior use of the rubric. This information will help the facilitator prepare for the norming process by anticipating the questions and concerns of individual raters.

Finally, facilitators need to internalize the steps of the rubric norming process. Understanding these steps helps facilitators prepare themselves for the sometimes difficult process of leading raters to consensus through norming. It also helps the facilitator prepare their raters. Ideally, facilitators will seek opportunities to practice the norming process prior to a high stakes or highly visible norming scenario. Through practice, facilitators will anticipate opportunities and challenges presented by a particular set of student work samples and a specific rubric. Facilitators will also be able to identify a number of procedural tasks that can be completed ahead of time, such as de-identifying student work samples, selecting student work samples to

use during norming practice, making copies of student work samples and rubrics for raters, planning for audio-visual needs, and devising ways of capturing rater scores for individual student work samples. Facilitators may also want to appoint an assistant to help with these tasks, as well as the “care and feeding” of raters during the norming session.

At the start of a norming session, facilitators can take charge and put raters at ease by welcoming them, initiating introductions, distributing materials including copies of student work samples, the rubric, any IRB or consent forms, nametags, and relevant documents such as learning outcomes, standards, or assignment descriptions. The facilitator should also share an outline or agenda for the norming session. Raters will appreciate having an overview of the steps of norming process and any special guidelines in advance.

Official Rule #1: Be a role model.

The first official step of the norming process is for the facilitator to think aloud through scoring several examples (Oakleaf, 2006).

In this step, the facilitator serves as a role model by explaining how s/he applied the rubric to student work samples. Facilitators can begin by asking the raters to read or view one student work sample silently and waiting quietly while they do so. Then the facilitator articulates how s/he scored the work sample using the rubric, asking the raters to hold questions until the explanation is complete. If the rubric is detailed and analytical, the facilitator can explain her/his scores methodically, going through one rubric criterion at a time. The facilitator may use phrases like, “I gave this work sample a ‘2’ or ‘Developing’ on the first criterion because of [A, B, and C],” or “I was tempted to give it a ‘1’ or ‘Beginning,’ but then I realized the student stated [X, Y, and Z].” The goal is to be descriptive and reveal the internal thought processes that resulted in a particular score. Once the “think aloud” is complete, the raters may ask questions. Depending on the length of student work samples, facilitators should think aloud through three to five examples. Ideally, these examples will represent the range of student responses: one very good, one very poor, and a few mid-range examples.

Unofficial Rule B: It’s not about you, facilitator.

While the facilitator serves as the leader of the norming session and begins by modeling the

rating thought process, facilitators must remember that the goal of norming is for the raters to come to consensus, not for the raters to agree with the facilitator. Good facilitators know that their own personal conceptions of the rubric are secondary to those belonging to the actual raters of student work. This means facilitators must walk a fine line: they need to lead the norming process, but not champion their own understanding of the rubric. They need to keep raters on track, but at the same time allow raters to take ownership of the rubric and the scoring process. This is particularly challenging for facilitators who were also involved in the production of the student work samples or development of the rubric. Facilitators might use sentences like, “To me, this student work sample aligns with [this specific language] in the rubric, but what is important is that you all agree as raters on how to handle this type of student response. How do you all see this fitting into the rubric?” or “How I understand this rubric criterion is [thus and so]. How can you all agree to interpret it?”

Official Rule #2: Let the raters try.

The second official step of the norming process is to ask raters to independently score a few student work samples (Maki, 2004).

After the think aloud stage, facilitators should guide raters through the scoring of a few work samples, one at a time. Ideally, these examples have been picked purposefully by the facilitator in advance so that they show an appropriate range of student responses. In this step, facilitators will want to use work samples that are straightforward, in that they can be scored easily and without too much confusion. For each work sample, facilitators can cue raters to read or view it quietly and independently. They may ask raters to consider silently, “What rubric score(s) would I assign this work sample?” Facilitators can answer questions about procedure during this time, but should defer discussions of scores until the next step. Then, when facilitators are finished with the first work sample, they can move on to the next. In this step, raters should score between one and three student work samples, depending on the length and complexity of the samples.

Official Rule #3: Take their temperature.

The third official step of the norming process is to bring raters together to determine how they scored the student work samples and look for patterns of consistent and inconsistent scores

(Maki, 2004).

After raters have scored a few samples independently, this step in the norming process allows raters to share their scores with the larger group. Facilitators should walk raters through each work sample, one at a time. For each work sample, facilitators should ask raters to share the scores they gave on each criterion of the rubric, one criterion at a time. For example, facilitators might ask, “How did you all score the first example on the first [criterion/row] of the rubric? or “For the first example, by a show of hands, how many of you gave it a ‘3’ or ‘Exemplary’ on the first rubric criterion? How many gave it a ‘2’? a ‘1’?” For this step, facilitators look for areas in which the raters agree and all assigned the same score. In these cases, facilitators can congratulate their raters on agreeing! Facilitators also identify areas in which the raters do not all agree and scores are not assigned consistently. In these cases, more practice and discussion is required.

Official Rule #4: Discuss and reconcile

The fourth official step of the norming process is to discuss and reconcile inconsistent scores (Maki, 2004).

At this stage, facilitators may ask “Where do we disagree? How can we come to consensus on how to score these student work samples?” Facilitators should review each rubric area where raters did not agree on a score. They may reread the rubric criterion aloud and then ask again for raters to share their scores as well as their rationale for assigning those scores. For example, facilitators may begin a conversation in this way: “Raters who assigned this a ‘2’, explain why you think a ‘2’ fits. Raters who assigned this a ‘1’, explain why you think a ‘1’ fits.” Once the rationales for these scores are clearly stated, facilitators can help raters move toward agreement. At this early stage, some disagreements can be reconciled, while others will remain.

Unofficial Rule C: The needs of the many outweigh the needs of the few, or the one.

During consensus-building discussions, facilitators may need to remind themselves that when it comes to norming a rubric, consensus is indeed the goal. For rubric assessments to be valid, they must first be reliable, and that means multiple raters must be able to provide consistent scores. During the early stages of rubric norming sessions, individual dissenting

opinions are valuable and should be surfaced and heard. Many times, what begins as an individual's disagreement with the larger group turns out to be a well-reasoned argument. In these cases, the group may reconsider its initial scores and change the way it rates future student work samples. Later in the norming process, individual dissent can be a difficulty that is a challenge to surmount, especially if the dissenting rater cannot either sway the opinion of larger group or decide to agree with the other raters.

Unofficial Rule D: It's okay to “disagree and commit.”

When an individual rater does not agree with the larger group, a facilitator may encourage that rater to “disagree and commit.” In this situation, the facilitator can review the need for consensus; s/he may emphasize the importance of coming to agreement through norming in order to establish inter-rater reliability. Then the facilitator may ask the rater to table their dissent and instead opt to agree to follow the rubric scoring practices of the group, at least temporarily. The facilitator might also offer to record the dissenting opinion so that it can be reviewed or incorporated at a later date.

Official Rule #5: Try again.

The fifth official step of the norming process is to repeat the independent scoring of a new set of student work samples (Maki, 2004).

After discussing the scores which raters gave the initial set of examples, it's time to try again. Facilitators should guide raters to score, independently and quietly, 1-3 additional student work samples, attempting to apply any new understandings from the previous discussion. In this step, facilitators may include work samples that are more challenging to score. Facilitators can circulate to answer any procedural questions that arise, but should hold full group conversation until all raters are ready to engage.

Official Rule #6: Separate wheat from chaff.

The sixth official step of the norming process is to bring all raters together again to review their scores/responses and identify patterns of consistency and inconsistency (Maki, 2004).

Facilitators can use the same questioning techniques described in steps 2 and 3 above to help raters share, in an organized way, the scores they assigned as well as their rationale for

assigning those scores. In this step, facilitators continue to identify and reinforce areas of agreement, while helping raters pinpoint and hash out areas of disagreement. At this point, facilitators may be able to determine which rubric areas are coalescing nicely (wheat) and which still need significant attention (chaff). In fact, if raters are in complete agreement about some areas, the facilitator may decide to focus exclusively on the problem areas during the rest of the norming session.

Sometimes, raters have difficulty coming to consensus on a particular rubric area because of problems that exist in the rubric itself. For example, raters may decide, as a group, to reject some content of the existing rubric and want to change it, or they may embrace some rubric content but dislike the language used and want to rephrase it. In some circumstances, the facilitator may have the latitude to revise the rubric so that raters can agree. If this is the case, facilitators should be prepared to help raters make changes that avoid common rubric mistakes and follow good rubric practices.

Unofficial Rule E: Stick to rubric best practices.

Although facilitators should not assert their own understanding of rubric content over a group of raters, they sometimes need to make suggestions related to good rubric practices. Facilitators who help raters revise rubrics in order to achieve consensus will want to steer raters away from common rubric mistakes (Oakleaf, 2009b). For example, raters sometimes suggest adding adjectives and adverbs to soften rubric language. However, the addition of words like “some,” “most,” “appropriately,” “clearly,” etc. expands the amount of subjective interpretation raters will apply to student work samples. As a result, this practice decreases rather than increases inter-rater reliability (Oakleaf, 2012). Another common mistake is to focus rubrics more on quantity of student performance (how many times a student exhibits a skill) rather than quality of student performance (how well the student exhibits a skill). A variation of this mistake is counting student errors (zero errors = mastery, 1-3 errors = developing, 4+ errors = novice); this does not measure the quality of student work, only the quantity of errors. When raters suggest changes to a rubric, a facilitator must always check the suggestions against a list of common mistakes to ensure the resulting change is aligned with good rubric practices. In these situations, the facilitator can play the role of the teacher,

providing guidance about alternative ways to improve a rubric.

Unofficial Rule F: Know when to hold ‘em, know when to fold ‘em.

In addition to advising raters about common mistakes and best practices, facilitators need to adopt an honest, direct, and candid tone when attempting to solve persistent rater agreement problems. Leading any group can be challenging, and leading one to agreement can seem unmanageable at times. Occasionally, a facilitator must help outspoken raters express their perspectives diplomatically or ensure that an introverted rater is heard. Sometimes tangential issues are raised repeatedly and a facilitator must gently but firmly place them on a “parking lot” list of issues to deal with at a later time. From time to time, raters may become overly negative about the assignments that generate student work samples or the student work samples themselves. In such cases, facilitators can acknowledge the comments, but then they must move the group past them and back to the task at hand. Often, humor is a useful tool.

Official Rule #7: Rinse and repeat.

The “final” official step of the norming process is to repeat the process “until raters reach consensus...ordinarily, two to three of these sessions calibrate raters’ responses” (Maki, 2004). Facilitators may choose to limit additional rounds to areas of the rubric that raters find difficult to agree on and omit the areas on which raters already agree.

Unofficial Rule G: End on a high note.

Before declaring a rubric “normed” and releasing raters to score student work samples independently, facilitators would do well to end the norming session in a positive way. Facilitators can congratulate raters on their persistence in seeking agreement, their exchange of ideas, their increased awareness of the complexities of understanding and analyzing student work samples, etc. If nothing else, they survived the norming process! As in most things, practice makes perfect, and the norming process itself can be rewarding even if there have been disagreements along the way.

Benefits of Norming

The benefits of norming a rubric are significant, but with so many steps and rules—official and unofficial—it's clearly a time-consuming process. Even with the recognition of the value of information literacy assessment and an understanding of rubrics and assessment best practices, there may be some skepticism, given the typical workload of busy academic librarians. Why is this norming process worthwhile? Why should we consider adding this process to our instruction and assessment practice on a regular basis?

From a teaching perspective, norming a rubric engenders rich discussions of instructional strategies and content that can improve information literacy programs. Rubric criteria that are easily agreed upon are usually areas of teaching and learning that are clear-cut. For example, a student either uses proper citation style or does not. There might be gradations of performance but not much to discuss philosophically among colleagues. However, where there are areas of heated discussion, there might also be quite a bit of discrepancy in how library instruction occurs for that particular criterion. There might be a wide range of instruction styles for a particular skill, and the norming experience might facilitate conversations among colleagues to share teaching experiences. Alternatively, library instructors might prioritize certain skills differently; in this case, norming might facilitate discussions about scope and sequence of library instruction. In general, the intense dialogue required to interpret rubric content can yield much personal reflection, group camaraderie, and eventual improvement in information literacy instruction.

From an assessment perspective, rubric norming also builds librarian assessment skills and confidence. In addition to strengthening their ability to construct and use rubrics, librarians gain experience that helps them think through future assessment designs and plan for additional collection and analysis of evidence of library impact on students.

Ultimately, norming rubrics leads to reliable and valid assessment—and using reliable and valid assessments is the key to producing believable and actionable results. Using a rubric to measure student learning, without giving consideration to the reliability and validity of that rubric, may be a waste of valuable time and energy. To ensure that the rubric and the data it generates are accepted by campus-wide stakeholders and usable for library decision-making,

norming is crucial. By following the official and unofficial rules for facilitating rubric norming sessions, librarians can perform this essential role and help demonstrate the value of academic libraries in the context of student success.

pre-print - see published version

References

- Gilchrist, D. & Oakleaf, M. (2012, April). *An essential partner: The librarian's role in student learning assessment*. Retrieved from http://www.learningoutcomeassessment.org/documents/LibraryLO_000.pdf
- Maki, P. L. (2004). *Assessing for learning: Building a sustainable commitment across the institution*. Sterling, VA: Stylus Publishing, LLC.
- Stevens, D. D., & Levi, A. (2013). *Introduction to rubrics: An assessment tool to save grading time, convey effective feedback, and promote student learning* (2nd ed.). Sterling, VA: Stylus Publishing, LLC.
- Oakleaf, M. (2006). *Assessing Information Literacy Skills: A Rubric Approach*. (Doctoral Dissertation). Retrieved from <http://meganoakleaf.info/oakleafdissertation.pdf>
- Oakleaf, M. (2009a). The information literacy instruction assessment cycle: A guide for increasing student learning and improving librarian instructional skills. *Journal of Documentation*, 65(4), 539-560. DOI 10.1108/00220410910970249
- Oakleaf, M. (2012). *RAILS 2010-2012 Rubrics*. Retrieved from <http://railsontack.info/media/documents/2012/6/railsIRR20102012.pdf>
- Oakleaf, M. (2009b). *Writing rubrics right: Avoiding common mistakes in rubric assessment*. Retrieved from <http://meganoakleaf.info/writingrubricsright.pdf>