

Norming a VALUE rubric to assess graduate information literacy skills

David J. Turbow, PhD; Julie Evener, MLIS

See end of article for authors' affiliations.

DOI: <http://dx.doi.org/10.3163/1536-5050.104.3.005>

Objective: The study evaluated whether a modified version of the information literacy Valid Assessment of Learning in Undergraduate Education (VALUE) rubric would be useful for assessing the information literacy skills of graduate health sciences students.

Methods: Through facilitated calibration workshops, an interdepartmental six-person team of librarians and faculty engaged in guided discussion about the meaning of the rubric criteria. They applied the rubric to score student work for a peer-review essay assignment in the "Information Literacy for Evidence-Based Practice" course. To determine inter-rater reliability, the raters participated in a follow-up exercise in which they independently applied the rubric to ten samples of work from a research project in the doctor of physical therapy program: the patient case report assignment.

Results: For the peer-review essay, a high level of consistency in scoring was achieved for the second workshop, with statistically significant intra-class correlation coefficients above 0.8 for 3 criteria: "Determine the extent of evidence needed," "Use evidence effectively to accomplish a specific purpose," and "Access the needed evidence." Participants concurred that the essay prompt and rubric criteria adequately discriminated the quality of student work for the peer-review essay assignment. When raters independently scored the patient case report assignment, inter-rater agreement was low and statistically insignificant for all rubric criteria ($\kappa = -0.16$, $p > 0.05$; $\kappa = 0.12$, $p > 0.05$).

Conclusions: While the peer-review essay assignment lent itself well to rubric calibration, scorers had a difficult time with the patient case report. Lack of familiarity among some raters with the specifics of the patient case report assignment and subject matter might have accounted for low inter-rater reliability. When norming, it is important to hold conversations about search strategies and expectations of performance. Overall, the authors found the rubric to be appropriate for assessing information literacy skills of graduate health sciences students.

Keywords: Information Literacy; Calibration; Educational Measurement; Education, Graduate; Interdepartmental Relations; Cooperative Behavior

Accrediting bodies hold institutions of higher education to high standards of accountability in measuring student learning. In striving to build and sustain cultures of assessment, an institution must emphasize systematically gathering evidence of progress at the programmatic and institutional levels. One common student learning outcome in

medical and health sciences education is "skills for lifelong learning," which connotes and encompasses information literacy. Studies show that some health sciences professionals lack these essential skills that they need for evidence-based practice [1]. Thus, the information literacy skills identified by the Association of College & Research Libraries (ACRL) [2] that are required for effective evidence-based practice make them particularly important for medical and health sciences students to acquire.



Supplemental Appendix and Appendix B are available with the online version of this journal.

One method for assessing information literacy skills is through rubrics. In the field of education, rubrics are standard methods to evaluate student performance. They usually have defined dimensions or characteristics of performance that can be measured (criteria). Evidence suggests that using analytic rubrics can be an effective way to examine learning outcomes related to information literacy [3, 4].

Rubric norming refers to the process in which workshops are conducted, with appropriate calibration activities so as to achieve a desired level of consensus about student performance criteria and standards of judgment—in other words, so that evaluation judgments are equally applied and fit the proposed student group [5]. While faculty and librarians can gather direct evidence of student learning with rubrics, they must take appropriate steps to ensure that rubrics are applied consistently and reliably across raters [4–6].

At a graduate health sciences university, the authors investigated using an information literacy rubric to track student progress in information literacy skills for various degree programs. We based the design of our information literacy rubric on the Association of American Colleges and Universities' (AAC&U's) information literacy Valid Assessment of Learning in Undergraduate Education (VALUE) rubric, which was developed by a national team of faculty who were content experts or closely involved in outcomes assessment. According to Finley, the VALUE rubrics have face validity and content validity [7]. Though the VALUE rubrics were designed to assess undergraduate learning, Gleason, Gaebelein, Grice, Crannage, Weck, Walter, and Duncan found a VALUE rubric to effectively track progression of critical thinking skills among graduate-level students [8]. Additionally, the information literacy VALUE rubric is heavily based on the ACRL "Information Literacy Competency Standards for Higher Education," which apply beyond undergraduate populations. We, therefore, decided that the structure of the information literacy VALUE rubric was appropriate for graduate students as an assessment instrument.

Our modification involved changing the word "information" to "evidence" throughout the rubric to more closely align the rubric to a health sciences curriculum. Our health sciences faculty members and students conceive of information literacy in an applied context of skill-based development of evidence-based practice, thus making "evidence" a more natural term for them to use.

The purpose of this project was to collaborate to test the utility of an information literacy rubric to assess the students' information literacy skills. The main goal of the project was to determine whether an interdepartmental team of raters, once trained, would find a modified version of the information literacy VALUE rubric to be appropriate for graduate level work in the health sciences. Specifically, the project addressed whether the design of the VALUE rubric discriminates quality of student work for research-based assignments. The project also addressed whether the language of the rubric, including the criteria and performance levels, facilitates calibration among raters.

METHOD

We piloted a version of the modified "Information Literacy Assessment Rubric," based on the AAC&U information literacy VALUE rubric [9]. The director of library services provided consultation on rubric criteria and definitions (Appendix A, online only). The university's institutional review board approved the project.

We embedded the rubric in the "Information Literacy for Evidence-Based Practice" course and two research courses. In the following trimester, we assembled a voluntary interdepartmental team of librarians and faculty from the doctor of physical therapy (DPT) and master of orthopaedic assistant (MOA) programs to serve as raters to calibrate the rubric. The outcomes assessment coordinator was self-selected to be the workshop facilitator. The director of library services, who teaches the information literacy course, selected three samples of student work on the peer-review essay assignment for rubric calibration. Student work was de-identified by name, student identification number, and session. One week before the first calibration workshop, the facilitator circulated the rubric and essay samples to participants. The facilitator then asked the team, comprising three faculty members and three librarians, to independently apply the rubric to score the samples.

First calibration workshop

At the first calibration workshop, the facilitator guided participants in a discussion about the meaning of each rubric criterion. Participants then reviewed an inter-rater summary table of their rubric scores. Participants discussed their impressions of

each sample of student work and their rationales for assigning scores on each criterion. The group sought consensus in scoring across raters. For each rubric criterion, the facilitator noted whether raters reached consensus on their scores along with any residual area of disagreement [5]. Shortly after the first calibration workshop, raters independently scored a second sample of de-identified student work from the same assignment.

Second calibration workshop

The facilitator presented a summary table of scores from the second exercise. Participants provided qualitative perceptions on whether there was heightened consistency in ratings after the first workshop. The group then repeated the consensus-seeking steps in rubric scoring and noted areas of disagreement.

For the second calibration workshop, inter-rater reliability of scoring was determined for each rubric criterion via intra-class correlations (ICCs) in a two-way mixed model. We considered our six raters to be a fixed effect and the three selected essays to be a random effect in the model.

Post-calibration activity

After the second rubric calibration workshop, inter-rater reliability of the rubric was determined through a follow-up exercise in which raters independently applied the rubric to assess performance on a third sample of work: a different assignment for a different course.

The researchers selected an excerpt from the patient case report, a project for students in the DPT program, for inter-rater reliability analysis. Raters received only the introduction section of the patient case reports, which included a literature review, along with a list of end references for the entire project. The faculty rater most familiar with the patient case report assignment selected thirty samples.

Each rater independently reviewed ten papers, and each sample of work was examined by exactly two raters. We arbitrarily separated the scores of our six raters into three pairs of librarian-faculty combinations for analysis. We used LiveText, an assessment management system, to gather post-calibration rubric scores on student performance and measures of central tendency.

Third workshop and debriefing session

Following independent scoring of the patient case report assignment, the facilitator presented an inter-rater scoring summary to workshop participants. Raters discussed the utility of the rubric. Participants revisited the criteria descriptors and skill descriptors and suggested changes to the rubric. Participants then discussed their experience in applying the rubric.

RESULTS

First calibration workshop

After discussing each rubric criterion and reviewing the ratings of each student's paper, raters expressed a sense of heightened clarity on the purpose and direction of the norming project itself. Raters then compared their scores. Informally, raters also shared that they felt a greater level of comfort in communicating with one another after the first workshop.

Librarians expressed that they interpreted and applied the rubric criteria more narrowly and specifically when scoring student work, whereas faculty rated student essays more inclusively on broader dimensions of content development, context, and level of professionalism.

The independent scores of raters for the peer-review essay in the information literacy course are shown in online only Appendix B.

A major area of disagreement among raters was for the criterion "Access the needed evidence," which might have been a confusing criterion descriptor. The interpretation of the librarians in the group was that they could score this criterion on the evidence furnished by the student in the essay, based on the assumption that good evidence emerges from sound search strategies and quality sources of evidence. Faculty in the group found it difficult to score the criterion in cases where an assignment did not require students to describe their search strategies outright. Because raters completed scoring before this discussion, one rater did not assign scores for the criterion "Access the needed evidence" because she felt it was not applicable to the assignment (Appendix B, online only).

Second calibration workshop

Congruence in scores for the peer-review essay assignment was high, with intra-class correlations

Rubric criterion	Intraclass correlation coefficient	95% C.I. lower bound	95% C.I. upper bound	F-ratio	df1	df2
Determine the extent of evidence needed	0.92*	0.54	0.99	11.92	2	10
Access the needed evidence	0.82†	-0.02	0.99	5.59	2	10
Evaluate evidence and its sources critically	0.78†	-0.18	0.99	4.62	2	10
Use evidence effectively to accomplish a specific purpose	0.83†	-0.09	0.99	5.98	2	10
Access and use evidence ethically and legally	0.44	-10.83	0.99	1.80	2	12

Note: Each row shows the intra-class correlation coefficient (ICC) of the corresponding rubric criterion. ICCs are absolute agreement values obtained from two-way mixed models.
 * $p < 0.01$. † $p < 0.05$.

Table 1
 Inter-rater reliability for scoring of the peer review essay in the second workshop

above 0.8 and statistically significant for 3 criteria: “Determine the extent of evidence needed,” “Use evidence effectively to accomplish a specific purpose,” and “Access the needed evidence.” The highest level of inter-rater reliability between raters was for the rubric criterion “Determine the extent of evidence needed,” where intra-class correlation was 0.92 (Table 1). A high degree of inter-rater reliability between raters was also found for the criteria “Use evidence effectively to accomplish a specific purpose,” with intra-class correlation at 0.83, and “Access the needed evidence,” with intra-class correlation 0.82. An acceptable level of inter-rater reliability between raters was found for the criterion “Evaluate evidence and its sources critically” at 0.78. The only rubric criterion for which inter-rater reliability would be considered low was “Access and use evidence ethically and legally,” where intra-class correlation coefficient was 0.44.

Raters expressed that the experience of scoring work on the peer-review essay was simplified by virtue of their participation in the initial rubric calibration exercise. When reviewing their own independent scores in a summary table, raters noted that qualitatively their scores were more congruent, compared to the first calibration workshop. The raw numeric rubric scores from the second rubric calibration workshop are presented in online only Appendix B.

Post-calibration inter-rater reliability

After the second calibration workshop, raters independently applied the rubric to assess performance for a third, larger sample of student work from the patient case report assignment (n=30). Each of the 6 raters independently reviewed 10

samples of de-identified student work. Exactly 2 raters independently scored each sample of work, allowing for calculation of a Cohen’s kappa statistic for each rubric criterion [10].

There was low inter-observer agreement for all rubric criteria for the patient case report assignment (Table 2). Raters agreed least on the criteria: “Access and use evidence ethically and legally” (kappa=-0.158, $p > 0.05$), “Determine the extent of evidence needed” (kappa=-0.08, $p > 0.05$), and “Access the needed evidence” (kappa=-0.025, $p > 0.05$), where association was negative. Similarly, there was very low agreement (kappa=0.024, $p > 0.05$) between independent ratings on “Evaluate evidence and its sources critically.” Of the 5 dimensions of information literacy that the rubric intended to measure, the highest level of inter-rater agreement was for the criterion “Use evidence effectively to accomplish a specific purpose” (kappa=0.118, $p > 0.05$).

DISCUSSION

While there was strong inter-observer agreement when raters independently applied the rubric to score the peer-review essay assignment, inter-rater reliability was low when raters were asked to apply the same rubric to excerpts from a different type of assignment, the patient case report.

Several factors likely influenced observed inter-rater reliability. First, there was a lack of familiarity among raters with the specifics of the patient case report assignment and the technical, physical therapy subject matter. Of the six raters, only two teach in the DPT program and have physical therapy specialty knowledge. Second, some raters expressed that the written guidelines for the patient case report assignment itself were unclear. Also, only the

Rubric criterion	Kappa	Asymptotic standard error*	Approximate T†	Approximate significance
Determine the extent of evidence needed	-0.077	0.118	-0.626	0.531
Access the needed evidence	-0.025	0.128	-0.195	0.845
Evaluate evidence and its sources critically	0.024	0.120	0.213	0.831
Use evidence effectively to accomplish a specific purpose	0.118	0.145	0.940	0.347
Access and use evidence ethically and legally	-0.158	0.079	-10.611	0.107

* Not assuming the null hypothesis.
† Using the asymptotic standard error assuming the null hypothesis.

Table 2

Inter-rater reliability for scoring of patient case report

introduction section with the full set of end references for the entire paper was provided to raters. During the third workshop, some raters assumed incorrectly that students were required to separately submit a literature review as a component of the patient case report assignment. In actuality, students were expected to review the literature as part of their introductions.

Given that students submitted their work on a topic of their choice, raters expressed that it can be difficult to discriminate information competence based only on an introductory excerpt of the assignment as evidence. Librarian raters in particular expressed that they likely interpreted the criteria more stringently and narrowly when scoring student work without considering other strengths of the work. Faculty who teach the course were more familiar not only with the context of the assignment and course, but also with the caliber of work of students across trimesters.

Workshop participants unanimously expressed that the peer-review essay assignment was suitable for determining information competency among health sciences students in post-professional programs, and further, that the AAC&U information literacy VALUE rubric is appropriate as a scoring tool for the peer-review essay assignment in the information literacy course.

In reconciling their scores, participants stated that applying the rubric allowed them to adequately discriminate the quality of work for the peer-review essay assignment. Our findings are consistent with McConnell [11] in that applying rubrics increases reliability in scoring under certain circumstances.

The works of Hoffman and LaBonte [12], Holmes and Oakleaf [4], and Gola, Ke, Creelman, and Vaillancourt [13] speak to the importance of interdepartmental collaboration in assessment

projects. We sought to develop a partnership between assessment personnel, faculty members, and librarians at our university. Valuable partnerships can potentially be forged between academic departments and health sciences librarians to enhance information literacy instruction through rubric norming.

Limitations

The study had several limitations. First, only a small sample of student work—six essays—was used for rubric calibration. Thus, there may have been low variability in the quality of work upon which the rubric was calibrated.

Second, because there was an insufficient number of available student samples from the peer-review essay in the information literacy course on which to establish post-calibration inter-rater reliability, a decision was made to instead use student samples from a different assignment in a different course, the patient case report, for this portion of the project. Due to the constraints of the project and the team's schedules, inter-rater reliability was calculated on a relatively small sample of thirty essays independently scored by six raters for sixty readings total. In a review article by McConnell [11], research on the use of rubrics reveals a high level of difficulty in achieving statistically appropriate levels of reliability.

Additionally, though this project involved an interdepartmental team, neither deans nor academic program directors were available to participate in norming the rubric. Toward this end, follow-up research on applying the modified rubric to coursework across multiple programs will require the participation of program directors and deans, as recommended by Allen [10]. Perceived strength of

partnerships and cooperation were not formally measured.

Finally, during the project, the ACRL published a new document, the “Framework for Information Literacy for Higher Education” [14], which they intend to eventually replace the “Information Literacy Competency Standards for Higher Education,” on which the AAC&U VALUE rubric was based. Currently, the two documents coexist, but in the future, a rubric that better reflects the framework may prove beneficial. Further research would be necessary to determine the effectiveness of that rubric.

DISCLOSURES

The authors declare that they have no competing interests. Research was performed with no external funding.

ACKNOWLEDGMENTS

The authors acknowledge their teammates on this collaborative project: Jessica Cain, Kathleen Hagy, Marilyn Miller, Michelle Rojas, and Margaret Wicinski. The supportive efforts of Patricia Mon and Bryce Pride are appreciated. Amy Driscoll provided important guidance and formative feedback on the project.

REFERENCES

1. Scurlock-Evans L, Upton P, Upton D. Evidence-based practice in physiotherapy: a systematic review of barriers, enablers and interventions. *Physiotherapy*. 2014 Sep; 100(3):208–19. DOI: <http://dx.doi.org/10.1016/j.physio.2014.03.001>.
2. Association of College & Research Libraries. Information literacy competency standards for higher education [Internet]. The Association; 18 Jan 2000 [cited 2 Jun 2015]. <<http://www.ala.org/acrl/standards/informationliteracycompetency>>.
3. Oakleaf M. Using rubrics to assess information literacy: an examination of methodology and interrater reliability. *J Am Soc Inf Sci Technol*. 2009 May;60(5):969–83. DOI: <http://dx.doi.org/10.1002/asi.21030>.
4. Holmes C, Oakleaf, M. The official (and unofficial) rules for norming rubrics successfully. *J Acad Librariansh*. 2013 Nov;39(6):599–602. DOI: <http://dx.doi.org/10.1016/j.acalib.2013.09.001>.

5. Maki P. *Assessing for learning: building a sustainable commitment across the institution*. Sterling, VA: Stylus Publishing; 2004.
6. Stevens D, Levi A. *Introduction to rubrics: an assessment tool to save grading time, convey effective feedback and promote student learning*. Sterling, VA: Stylus Publishing; 2004.
7. Finley AP. How reliable are the VALUE rubrics? *Peer Rev*. 2011/2012 Fall/Winter;13/14(4/1):31–3.
8. Gleason BL, Gaebelein CJ, Grice GR, Crannage AJ, Weck MA, Walter B, Duncan W. Assessment of students’ critical-thinking and problem-solving abilities across a 6-year doctor of pharmacy program. *Am J Pharm Educ*. 2013 Oct;77(8):1–12. DOI: <http://dx.doi.org/10.5688/ajpe778166>.
9. Rhodes T, ed. *Assessing outcomes and improving achievement: tips and tools for using rubrics*. Washington, DC: Association of American Colleges and Universities; 2010.
10. Allen M. *Assessing academic programs in higher education*. San Francisco, CA: Jossey-Bass; 2004.
11. McConnell KD. Rubrics as catalysts for collaboration: a modest proposal. *Eur J High Educ*. 2013;3(1):74–88. DOI: <http://dx.doi.org/10.1080/21568235.2013.778043>.
12. Hoffman D, LaBonte K. Meeting information literacy outcomes: partnering with faculty to create effective information literacy assessment. *J Inf Lit*. 2012;6(2):70–85. DOI: <http://dx.doi.org/10.11645/6.2.1615>.
13. Gola CH, Ke I, Creelman KM, Vaillancourt SP. Developing an information literacy assessment rubric: a case study of collaboration, process, and outcomes. *Communs Inf Lit*. 2014;8(1):131–44.
14. Association of College & Research Libraries. *Framework for information literacy for higher education* [Internet]. The Association; 2 Feb 2015 [cited 13 Oct 2015]. <<http://www.ala.org/acrl/standards/ilframework>>.

AUTHORS’ AFFILIATIONS



David J. Turbow, PhD, dturbow@usa.edu, Outcomes Assessment Coordinator, University of St. Augustine for Health Sciences, 700 Windy Point Drive, San Marcos, CA 92069; **Julie Evener, MLIS**, jevener@usa.edu, Director of Library Services, University of St. Augustine for Health Sciences, 1 University Boulevard, St. Augustine, FL 32086

Received November 2015; accepted February 2016