

Using Quantitative Research Methods to Determine the Validity and Reliability of an Undergraduate Citation Rubric

Katelyn Angell

Long Island University, Brooklyn, NY, USA

Abstract. Since the beginning of the 2000s academic librarians have been developing and using scoring rubrics to assess various facets of student learning within the library instruction classroom. There are numerous advantages to evaluating student learning outcomes with rubrics. These benefits include the capability to evaluate student work over time or across academic subjects, a method of offering students constructive, objective feedback, and an opportunity for library colleagues to create a unified, standardized evaluation tool. Currently there are many articles within Library and Information Science literature detailing the adoption of rubrics into the academic library classroom. However, there is little research within the field pertaining to reliability and validity analyses of these rubrics. In order to ensure that a rubric consistently measures student learning comprehension it is important to perform these quantitative calculations on it prior to widespread use. The present study describes a project the author completed to measure reliability and validity of a rubric developed to assess the citing and referencing knowledge of undergraduate students at a large university in New York City.

Keywords. Assessment, rubrics, validity, reliability, citing, referencing

1. Introduction

In recent years, academic librarians across the world have been increasingly incorporating scoring rubrics into their information literacy assessment processes. Moskal (2000) defined rubrics as “descriptive scoring schemes that are developed by teachers or other evaluators to guide the analysis of the products or processes of students’ efforts (Brookhart, 1999),” and can be used when “a judgment of quality is required and may be used to evaluate a broad range of subjects and activities” (2000). As the role of reference and instruction librarians continues to expand into classroom teaching it is integral that members of the profession devise standardized and valid tools for evaluating

student performance on information literacy tasks. In order for librarians to improve student learning outcomes it is vital to create means of tracking and assessing these outcomes.

One area in particular meriting assessment among college students is citing and referencing. Years of professional experience combined with corroborating LIS scholarship supports the claim that properly documenting sources is difficult for many undergraduates. Head and Eisenberg (2010) surveyed over 8000 college students to learn about their information-seeking processes, finding that 41% of participants experienced difficulty with citing sources in their research papers. Additionally, 29% of students weren't even certain when citations are required (p. 25). These statistics illustrate that nearly half of college students within the United States have trouble citing their sources. Given the importance of documentation to academic writing strategies must be undertaken to help students improve their citing abilities.

In addition to teaching these skills in the classroom and disseminating practical activities librarians might also want to consider directly evaluating student knowledge of citing and referencing. One method of accomplishing this goal would be to create a citing skills worksheet and a rubric to grade these assignments. In order for the rubric to be most effective, it is important that its reliability and validity are assessed prior to widespread use.

2. Literature Review

While the bulk of research on rubrics within Library and Information Science (LIS) literature has occurred in the 2000s, education scholars have been employing and documenting the use of rubrics since the 1990s. Cohen (1995) described a pioneering project in which college freshmen and teaching faculty worked collaboratively to develop reflection and coursework rubrics. These rubrics were used not solely by the professor, but were also used by students to assess their own work and that of their classmates.

Currently, rubrics are frequently used by both school and academic librarians to evaluate student progress on library-related tasks. There are many tangible benefits to using rubrics as assessment tools within the information literacy classroom. Nichols, writing in an article compiled by Fontichiaro (2011), elucidates two advantages of rubrics, writing that they "show our students where they rank in demonstrated experience in that area," and can also provide "detailed feedback without extended grading time" (p. 12). Therefore, rubrics offer students clear and objective feedback on their strengths and weaknesses on a particular learning object. The format of a rubric, which outlines specific criteria and performance levels, as shown in Figure 1, streamlines teacher expectations and grading details. An additional benefit of rubrics is that it provides librarians with a forum for collaborating to develop collective and consistent information literacy objectives and standards (Van Helvoort, 2010).

There is increasing evidence within LIS that rubrics are an authentic means of measuring student learning outcomes. Over the past fifteen years many academic librarians have documented the successful application of rubrics to information literacy practices. These rubrics are used to measure many competencies associated with information literacy among students at academic institutions. Choinski, Mark, and Murphey (2003) used the reflection papers of students enrolled in credit-bearing information services classes to assess learning outcomes. They used a detailed rubric which evaluated many important components of information literacy, such as identifying and finding library resources, evaluating sources, and understanding differences between popular and scholarly sources.

Knight (2006) used a rubric to score annotated bibliographies produced by first year students. Five popular information literacy learning objectives were included on this rubric, such as “evaluates credibility of information” and “locates scholarly books and articles, reference works, [and] credible web sites” (p. 47). Boss and Drabinski (2014) used a set of rubric-based content analysis questions to analyze the syllabi of undergraduate and graduate business classes. Their objective was to better embed programmatic information literacy instruction within this department. While the bulk of this literature concerns undergraduate populations there has also been some research on using rubrics with graduate students. For example, librarians and teaching faculty at a university in Virginia partnered to develop a rubric intended to evaluate the literature reviews of graduate student theses (Green & Bowser, 2006).

One important information literacy competency warranting additional assessment research is referencing and citing. Knight (2006) included “recognizes plagiarism and formats citations in correct style” (p. 47) as one of the criteria in an information literacy rubric mapped to the Association of College and Research Libraries’ (ACRL) Information Literacy Competency Standards for Higher Education (2000). Analysis in Knight’s study revealed that a large number of the student participants “failed to use a consistent and correct citation format, relying instead upon the style of notation employed by the database” (p. 50). Citing is also often anecdotally discussed between librarians and teaching faculty as a problem area for students, and many members of both groups brainstorm creative methods of helping students to improve these skills. Kargbo (2010) conducted valuable research exploring the particulars of documentation troubles among undergraduates. A sample of 675 college students was surveyed to determine both their citing abilities and their confidence in these abilities. Unfortunately, the majority of participants reported that they were not confident with their citing capabilities (62%). One salient cause of student confusion was identified as the massive amounts of citation styles available. Another issue was learning how to distinguish between various formats, such as the differences in citing a book versus an academic journal article. Many suggestions to remedy this problem were put forth, such as

providing students with guides and adopting a uniform citation style across departments.

Although there is increasing support in LIS and educational literature for rubrics as a valuable assessment tool, the rubrics themselves need to be evaluated prior to widespread application. Creating and using a rubric without assessing factors such as reliability and validity can potentially lead to ineffective grading of student artifacts (Rezaei & Lovorn, 2010). According to Reddy and Andrade (2010), the effectiveness of a rubric can be evaluated “by determining if it measures what it is intended to measure (validity) and provides for consistency in scoring (reliability)” (p. 441). There are many different methods of calculating reliability and validity for testing instruments. One popular reliability measure for rubrics is interrater reliability, used by Bresciani and collaborators (2009) to determine if all prospective users of a grading rubric for college students were consistent in their judgments. Another LIS researcher using interrater reliability within rubric assessment is Oakleaf, with two articles published in 2009 on this topic.

Internal consistency is a common reliability measurement used within research on educational assessment tools. The objective of internal consistency measures is to determine if the separate items on an instrument are significantly related. “If applied to a rubric, the goal of an internal consistency measure would be to determine whether all the different dimensions of the rubric measure the same construct” (Angell, 2015, p. 4). Cronbach’s alpha is a statistic often employed by social science researchers to measure reliability. At this point in time there is little mention of reliability analysis in LIS literature (Walsh, 2009).

In addition to reliability, it is important to investigate the validity of an assessment tool. Knowledge that an instrument is valid assures researchers that that they are measuring the construct they intended to measure. A thorough literature review by Walsh (2009) revealed that that only a few LIS researchers conducted validity analyses on their information literacy testing instruments. Although there are many ways to measure validity one measure particularly applicable to testing instrument evaluation is content validity. The latter can be defined as “the degree to which the elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose” (Haynes, Richard, & Kubany, 1995, p. 238). Experts on a particular topic, sometimes referred to as “subject matter experts,” (SMEs) are generally used to decide if a rubric has content validity. Jonsson and Svingby (2007) examined 75 studies featuring scoring rubrics to determine the frequency of validity and reliability calculations. The target populations were varied, ranging from primary schools to colleges and universities. Ten of these studies incorporated content validity tactics into their procedures for assessing scoring rubrics. Green and Bowser (2006) are cited in this robust review, but theirs appears the only study pertaining to information literacy instruction. At this point in time content validity has only been documented a few times in LIS

literature. Erfanmanesh, Abrizah, and Karim (2012) used seven LIS subject matter experts to validate the Information Seeking Anxiety Scale. This process resulted in the researchers acquiring evidence that their scale would appropriately measure the construct in question. Another study incorporating content validity was conducted by Clark and Catts (2007). Both subject matter experts and medical students were consulted to determine if the items on an information literacy skills inventory were satisfactory or needed alterations.

One discipline which uses innovative content validity measurements on assessment instruments is nursing (Polit & Beck, 2006). One particular method employed by nursing scholars, the Content Validity Index (CVI), holds promising potential for information literacy assessment tools. The CVI allows subject matter experts to evaluate the validity of each test item and the instrument as a whole.

The prevalence of LIS articles detailing the use of rubrics combined with a lack of articles documenting reliability and validity measurements necessitates greater research into this burgeoning area. The primary aim of this project was to develop a rubric assessing undergraduate mastery of a difficult information literacy competency, citing and referencing, and use internal consistency and content validity measurements to evaluate its reliability and validity. These measures help ensure that the rubric is effective and objective in meeting its goal of improving student learning outcomes.

3. Method

3.1. Participants and Materials

There were two distinct participant groups in this study. The first were two classes of 42 undergraduates enrolled in an interdisciplinary social sciences class at a large urban university in New York City. Twenty-five students were first years and 17 were sophomores. All students visited the library at two points during the semester for information literacy instruction. The second group of participants was 42 librarians recruited through professional listservs.

Prior to working with the participants the researcher created a rubric to assess the documentation skills of undergraduates. Both the American Psychology Association (APA) and Modern Language Association (MLA) styles were included (see Figure 1). The rubric was used to grade a four question in-class activity. The four questions on the student assignment were:

1. Find an article in a sociology database on your topic and email it to yourself and the librarian.
2. Using APA style, write down the citation for the article you found.
3. Using MLA style, write down the citation for the article you found.
4. List one difference you notice between APA and MLA citation styles.

In order to evaluate the rubric's content validity a four-item survey was created in Google Forms. The survey asked the 42 librarians (subject matter experts) to rate the relevance of the rubric's four criteria on a Likert scale from one (irrelevant) to four (extremely relevant). This format was designed using the Content Validity Index as a guide.

Figure 1. Original Citation Rubric¹

Criteria	Level 1 Beginning	Level 2 Developing	Level 3 Advanced
Locating a journal article in a specified database and emailing it to the instructor	Does not email a journal article	Emails a journal article from an incorrect database	Emails a journal article from the specified database
Citing an electronic journal article in APA (6 th edition) format	Three or more errors in formatting article in APA or leaves question blank	One or two errors in formatting article in APA format	Cites article in perfect APA format
Citing an electronic journal article in MLA (7 th edition) format	Three or more errors in formatting article in MLA or leaves question blank	One or two errors in formatting article in MLA format	Cites article in perfect MLA format
Listing differences between MLA and APA styles	Provides incorrect answer or leaves question blank	Provides vague or partially incorrect answer	Correctly lists one or more differences between MLA and APA styles

3.2. Procedure

The content validity survey was distributed to other librarians via a popular listserv, as well as promoted via word of mouth. All responses were anonymous. Once all surveys were completed the content validity was calculated at the item level (I-CVI) and for the rubric as a whole (S-CVI). The CVI is assessed by adding up all of the threes and fours for each item rated by the subject matter experts and then dividing the sum by the total number of surveys. Acceptable

¹ Reprinted with permission from Angell's (2015) article in *Behavioral & Social Sciences Librarian*

levels of I-CVI with six or more raters is >0.78 and >0.8 for S-CVI (Polit & Beck, 2006).

Secondly, all 42 in-class assignments were graded using the citation rubric. Afterward, internal consistency reliability was calculated for the citation rubric using Cronbach's alpha.

IV. Results

First, the subject matter expert ratings of the relevance of the four rubric criteria were averaged. The lowest rating option was one (irrelevant) and the highest was four (very relevant). Collectively, the librarian participants ranked the four criteria from least to most significant as follows: locating a journal article (2.71), listing a difference between MLA and APA styles, (2.89), citing an article in MLA (3.64), and citing an article in APA (3.67).

Next, the CVI was used to assess the rubric at both the item level and holistically. All ratings of three or four were added up and divided by the total number of responses. This data is presented in Table 1.

Table 1. CVI Scores as Rated by Subject Matter Experts

	Locate article	APA	MLA	Difference	Rubric Total
CVI	0.60	0.90	0.88	0.60	0.76

Two of the four rubric criteria, citing a journal article in MLA and APA styles, scored above the acceptable CVI level of 0.78, while the other two tasks did not meet the cutoff. The S-CVI was 0.76, just barely failing to reach the acceptable level.

In addition to validity, the internal consistency of the rubric was assessed using Cronbach's alpha. All 42 student citation activities were graded using the citation rubric. The scale ranged from one (beginning) to three (advanced). The lowest overall scores were for citing in APA (0.62), followed by locating an article (1.05), citing in MLA (1.07), and listing a difference between the two styles (1.62).

Data analysis of the scores yielded a Cronbach's alpha level of 0.51, which falls below the accepted level of 0.7 (Bresciani et al., 2009). Item-Total statistics for the rubric were also examined to evaluate the rubric's reliability.

Table 2. Citation Rubric Item-Total Statistics

	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Email article	.18	.59
Citing in APA	.39	.40
Citing in MLA	.49	.24

Difference between APA & MLA	.24	.49
------------------------------	-----	-----

The Corrected Item-Total Correlation (CITC) is used to determine correlations between each of the four rubric criteria and the overall score. Correlations falling below the level of acceptance, 0.3 (de Vaus, 2013), are undesirable because they suggest that the criteria in question fails to fit in with the rubric as a whole. In this case, the email article and difference between styles criteria do not meet the minimum acceptable level.

Next, the Cronbach's Alpha if Item Deleted (CAID) column states what the alpha level of the rubric would be if one of the criteria was deleted. Because the citation rubric received a Cronbach's alpha level of 0.51, the only criterion to take into consideration is the email article task, as it's the only criteria which would result in a higher alpha level if eliminated.

4. Discussion

The reliability and validity calculations applied to the citation rubric indicated that the instrument could benefit from some alterations. Data generated from the CVI suggested that two of the rubric's criteria, emailing an article and listing a difference between MLA and APA styles, were not integral parts of the rubric's overarching goal, to assess student citing and referencing skill sets. However, the subject matter experts showed strong support for retaining the criteria requesting students to cite a journal article in MLA (I-CVI: 0.88) and APA (I-CVI: 0.90) styles.

In addition, the relatively low Cronbach's alpha level of the rubric, 0.51, signified that its four tasks lacked a close enough association to warrant further use as a classroom assessment tool. The CITC data presented in Table 2 suggest that the email article and listing a difference tasks are not valid measurements of undergraduate APA and MLA citing and referencing knowledge. Bolstering this viewpoint is the CAID data, which shows that the rubric's overall internal consistency would improve without the email task. In contrast, the elimination of either the citing in MLA or APA task would cause the alpha level to drop dramatically, illustrating the importance of these two criteria to the rubric's effectiveness as a citing assessment tool.

The reliability and validity measurements prompted a major revision of the rubric (see Figure 2). The email article and listing difference tasks were deleted and replaced with two new criteria, which asks students to format in-text citations for a journal article in MLA and APA styles. The addition of these new rubric criteria were inspired by the high ratings of the librarian subject matter experts for the original two MLA and APA tasks. All future citation activities distributed to students will be graded with the new rubric. Hopefully, future

assessments of the modified rubric would result in higher content validity scores from subject matter experts.

Figure 2. Modified Citation Rubric²

Criteria	Level 1 Beginning	Level 2 Developing	Level 3 Advanced
Format an in-text citation in APA (6 th edition) for an electronic journal article	Two or more errors in formatting in-text citation or leaves question blank	One error in formatting in-text-citation	Cites article in perfect APA format
Cite an electronic journal article in APA (6 th edition) for References page	Three or more errors in formatting article in APA or leaves question blank	One or two errors in formatting article in APA format	Cites article in perfect APA format
Format an in-text citation in MLA (7 th edition) for an electronic journal article	Two or more errors in formatting in-text citation or leaves question blank	One error in formatting in-text-citation	Cites article in perfect MLA format
Cite an electronic journal article in MLA (7 th edition) for Works Cited page	Three or more errors in formatting article in MLA or leaves question blank	One or two errors in formatting article in MLA format	Cites article in perfect MLA format

There are two salient limits associated with the present study. To begin with, APA and MLA are just two out of many citation styles used by college students. The former is primarily used within the social sciences and the latter within the humanities. It would be beneficial for librarians to conduct a similar study with other popular citation styles, such as Chicago, Harvard, or the Council of Science Editors (CSE).

Another limitation of this study is that its main focus was assessing the effectiveness of a citation rubric, not evaluating student citing and referencing competencies. The researcher hopes to embark upon a future project which would use the modified citation rubric to formally evaluate student citing

² Reprinted with permission from Angell's (2015) article in *Behavioral & Social Sciences Librarian*

knowledge, and use this data to strategize methods for enhancing student learning comprehension of this important component of information literacy. Before librarians can make conclusions about student information literacy abilities it is pivotal to ensure that the instruments used to measure these competencies are both reliable and valid.

References

- American Library Association, (2000). *Information Literacy Competency Standards for Higher Education*. Available at: <http://www.ala.org/acrl/sites/ala.org.acrl/files/content/standards/standards.pdf>
- Angell, K., (2015). The application of reliability and validity measures to assess the effectiveness of an undergraduate citation rubric, *Behavioral & Social Sciences Librarian*, Vol. 34, No. 1, 2-15.
- Boss, K., & Drabinski, E., (2014). Evidence-based instruction integration: A syllabus analysis project, *Reference Services Review*, Vol. 42, No. 2, 263-276.
- Bresciani, M.J., Oakleaf, M., Kolkhorst, F., Nebeker, Barlow, J., Duncan, K., & Hickmott, J., (2009). Examining design and inter-rater reliability of a rubric measuring research quality across multiple disciplines, *Practical Assessment, Research & Evaluation*, Vol. 14, No. 12, 1-7.
- Choinski, E., Mark, A. E., & Murphey, M., (2003). Assessment with rubrics: An efficient and objective means of assessing student outcomes in an information resources class, *Portal : Libraries and the Academy*, Vol. 3, No. 4, 563-575.
- Clark, C., & Catts, R., (2007). Information Skills Survey: Its application to a medical course, *Evidence Based Library and Information Practice*, Vol. 2, No. 3, 3-26.
- Cohen, C., (1995). *Improving the freshman college classroom through building a purposeful community of altruistic and motivated learners*. Available at: <http://eric.ed.gov/?id=ED381216>
- de Vaus, D., (2013). *Surveys in Social Research*. Routledge, New York, New York.
- Erfanmanesh, M., Abrizah, A., & Karim, N.H.A., (2012). Development and validation of the Information Seeking Anxiety Scale, *Malaysian Journal of Library & Information Science*, Vol. 17, No. 1, 21-39.
- Fontichiaro, K., (2011). Nudging toward inquiry: Summative assessment, *School Library Monthly*, Vol. 27, No. 7, 12-13.
- Green, R., & Bowser, M., (2006). Observations from the field: Sharing a literature review rubric, *Journal of Library Administration*, Vol. 45, No. 1/2, 185-202.
- Haynes, S. N., Richard, D. S., & Kubany, E. S., (1995). Content validity in psychological assessment: A functional approach to concepts and methods, *Psychological Assessment*, Vol. 7, No. 3, 238-247.
- Head, A.J., & Eisenberg, M.B., (2010). *How College Students Evaluate and Use Information in the Digital Age*. Available at: <http://files.eric.ed.gov/fulltext/ED535166.pdf>
- Jonsson, A., & Svingby, G., (2007). The use of scoring rubrics: Reliability, validity and educational consequences, *Educational Research Review*, Vol. 2, 130-144.
- Kargbo, J. A., (2010). Undergraduate students' problems with citing references, *Reference Librarian*, Vol. 51, No. 3, 222-236.
- Knight, L. A., (2006). Using rubrics to assess information literacy, *Reference Services Review*, Vol. 34, No. 1, 43-55.
- Moskal, B.M., (2000). Scoring rubrics: What, when and how?, *Practical Assessment, Research & Evaluation*, Vol. 7, No. 3. Available at: <http://pareonline.net/getvn.asp?v=7&n=3>

Oakleaf, M., (2009). The information literacy instruction assessment cycle: A guide for increasing student learning and improving librarian instructional skills, *Journal of Documentation*, Vol. 65, No. 4, 539-560.

Oakleaf, M., (2009). Using rubrics to assess information literacy: An examination of methodology and interrater reliability *Journal of the American Society for Information Science and Technology*, Vol. 60, No. 5, 969-983.

Polit, D.F, & Beck, C.T., (2006). The Content Validity Index: Are you sure you know what's being reported? Critique and recommendations, *Research in Nursing & Health*, Vol. 29, No. 5, 489-97.

Reddy, Y.M, & Andrade, H., (2010). A review of rubric use in higher education, *Assessment & Evaluation in Higher Education*, Vol. 35, No. 4, 435-448.

Rezaei, A. R., & Lovorn, M., (2010). Reliability and validity of rubrics for assessment through writing, *Assessing Writing*, Vol. 15, No. 1, 18-39.

Van Helvoort, J., (2010). A scoring rubric for performance assessment of information literacy in Dutch higher education, *Journal of Information Literacy*, Vol. 4, No. 1, 22-39.

Walsh, A., (2009). Information literacy assessment: Where do we start? *Journal of Librarianship and Information Science*, Vol. 41, No. 1, 19-28.