Using Rubrics to Collect Evidence for Decision-Making:

What do Librarians Need to Learn?

Introduction

Every day, librarians make decisions that impact the provision of library products and services. To formulate good decisions, librarians must be equipped with reliable and valid data. Unfortunately, many library processes generate vast quantities of unwieldy information that is ill-suited for the evidence based decision-making (EBDM) practices librarians strive to employ. As a result, librarians require tools that facilitate the translation of unmanageable facts and figures into data that can be used to support decision-making. One such tool is a rubric.

Rubrics are "descriptive scoring schemes" used to analyze (Moskal, 2000) and judge the quality of services, products, or performances (Popham, 2003, p. 95). In libraries, rubrics can be employed to examine and evaluate a multitude of library products and services. For example, rubrics can capture useable data about information-seeking behavior, customer service skills, marketing and outreach efforts, collection strengths, and information commons effectiveness. This paper reviews the benefits of using rubrics to facilitate EBDM and describes a study in which librarians applied a rubric to capture and understand evidence of student information literacy skills—evidence that led to the improvement of a library instruction tutorial. This paper also identifies difficulties librarians may encounter when using rubrics and recommends training topics to maximize the usefulness of a rubric approach to EBDM.

Definition and Benefits of Rubrics

Rubrics are tools that describe the parts and levels of performance of a particular task, product, or service (Hafner, 2003, p. 1509). Rubrics are often employed to judge quality (Popham, 2003, p. 95) and they can be used across a broad range of subjects (Moskal, 2000). Full model rubrics are the most descriptive type of rubric. Formatted on a grid or table, full model rubrics include target indicators or criteria listed down the left hand side and levels of performance listed across the top of the grid (Callison, 2000, p. 34). The first of two components that comprise a full model rubric is criteria. Criteria are "the conditions a performance must meet to be successful" (Wiggins, 1996). Performance descriptors are the second component of rubrics. Performance descriptors

"spell out what is needed, with respect to each evaluative criterion…[for] a high rating versus a low rating" (Popham, 2003, p. 96).

Rubrics can be described as holistic or analytic. A holistic rubric "score[s] the overall process or product as a whole, without judging the component parts separately" (Nitko, 1996, p. 226). It gives one score for a whole product or performance based on an overall impression. Analytic rubrics "divide…a product or performance into essential traits or dimensions so that they can be judged separately—one analyzes a product or performance for essential traits. A separate score is provided for each trait" (Arter & McTighe, 2000, p. 18). To obtain a summary score from an analytic rubric, individual scores can be summed to form a total score (Nitko, 2000, p. 226). The rubric used in this study is a full-model, analytic rubric (see Figure 1).

Evaluation Criteria	Beginning	Developing	Exemplary
Articulates Criteria	0 - Student does not address authority issues.	1 - Student addresses authority issues, but does not use criteria terminology.	2 - Student addresses authority issues and uses criteria terminology such as: author, authority, authorship, or sponsorship.
Cites Indicators of Criteria	0 - Student does not address authority indicators.	1 - Student refers vaguely or broadly to authority indicators, but does not cite specific indicators.	2 - Student cites specific authority indicators such as: domain, server/publisher/host, or ~ in URL; presence of personal or corporate author name, email, "About Us" or "Contact Us" links; or author credentials.
Links Indicators to Examples from Source	0 - Student does not address examples of authority indicators from the site.	1 - Student refers vaguely or broadly to examples of authority indicators from the site under consideration, but does not cite specific examples.	2 - Student cites specific examples of authority indicators from the site under consideration.
Judges Whether or Not To Use Source	0 - Student does not indicate whether or not the site is appropriate to use for the purpose at hand.	1 - Student indicates whether or not the site is appropriate to use for the purpose at hand, but does not provide a rationale for that decision that cites authority issues and/or indicators.	2 - Student indicates whether or not the site is appropriate to use for the purpose at hand and provides a rationale for that decision citing authority issues and/or indicators.

RESEARCHER USE ONLY: Total Score ___/8

Figure 1. Study Rubric Assessing Student Evaluation of Website Authority

As a tool for EBDM, rubrics offer a number of benefits. First, rubrics provide librarians the opportunity to discuss, determine, and communicate agreed upon values (Callison, 2000, p. 36). Librarians who create rubrics must agree upon the criteria by which a library product or service will be analyzed and evaluated. They must also come to consensus about what the different performance levels of each criterion "look like". Thus, through the rubric development process, librarians engage in meaningful

discussions that result in shared views of the library product or service under scrutiny. Librarians who use existing rubrics as tools for EBDM can obtain similar benefits by engaging in activities and discussions to adapt a rubric to local needs. Once created or adapted, rubrics not only represent consensus views of librarians, they also communicate shared values to others, including new librarians or external stakeholders (Stevens & Levi, 2005, p. 23). Finally, they help "combat…accusations that evaluators do not know what they are looking for" (Bresciani et al., 2004, p. 30).

Rubrics offer a second important benefit for EBDM: descriptive, yet easily digestible data (Bresciani et al., 2004, p. 30). Because full model rubrics include descriptions of the key components of a library product or service at a number of performance levels, they provide detailed information for decision-making. At the same time, rubric data is easily simplified. For example, one might report that a library service operated at a "good" level during 80% of observed performances. If more detail is required, the rubric definition of "good" can be added to the percentage to make it more meaningful. In this way, rubric data can be reported in simple percentages or with details described in the rubric as the situation and audience require.

Rubrics offer librarians engaged in EBDM a third significant benefit. Because rubrics analyze agreed upon, detailed descriptions of library activities, they prevent inaccuracy of scoring (Popham, 2003, p. 95) and bias (Bresciani et al., 2004, p. 31). Rubrics clarify schemes for evaluation ahead of time, and therefore reduce subjectivity (Moskal, 2000). Since rubrics guide librarians to focus on essential criteria (Callison, 2000, p. 35), they can assess products and services more easily and objectively (Bernier, 2004, p. 25). Even in team or collaborative environments, rubric evaluations are "likely to be reasonably objective and consistent" (Callison, 2000, p. 35).

When rubrics are employed to make evidence based decisions about library instructional programs, as in the study described below, additional benefits are realized. Rubrics, more than many other educational assessment approaches, offer direct benefits to students. Because rubrics reveal the expectations of instructors and librarians, students can focus their efforts on achieving, rather than deciphering, learning goals. Because full model rubrics offer both numerical scores and descriptions of performance levels, ratings are more meaningful than letter or numerical scores alone (Bresciani et al., 2004, p. 35)

and provide students with feedback about what they have already learned and what they have yet to learn. As a result, rubrics support not only student learning, but also self-evaluation and metacognition.

Rubrics provide at least four major benefits to librarians seeking to use EBDM strategies and merit further investigation. To this end, this study examined 1) librarians' ability to use rubrics as a decision facilitation tool, 2) barriers that might prevent effective rubric usage, and 3) training topics that address potential barriers.

Methodology

This study investigated librarians' use of rubrics as an EBDM tool to improve an online information literacy tutorial. The data for the study came from student responses to open-ended questions embedded in an online information literacy tutorial called LOBO used by first-year students in English 101 at North Carolina State University (NCSU). Study participants applied rubrics to transform students' textual responses into quantitative data; this data was statistically analyzed for reliability and validity. Later, data that was determined to be valid was put to use for instructional decision-making.

Study Participants

Twenty-five participants joined this study. The participants were evenly divided into five groups: NCSU librarians, NCSU English 101 instructors, NCSU English 101 students, instruction librarians from other Association of Research Libraries (ARL) libraries, and reference librarians who had some instruction responsibilities from other ARL libraries. Librarian participants were selected in an attempt to represent a cross section of reference and instruction librarians in gender, race, and areas of expertise. English instructors and students were included for comparative purposes.

The internal (NCSU) participants (librarians, instructors, and students) took part in a rubric training session. Because most of the participants had little prior experience with rubrics, the researcher introduced rubrics by providing a definition, describing the component parts of rubrics, and providing the study rubric. The researcher followed a multi-step process to familiarize the participants with the task of scoring student responses. The researcher began by sharing five "anchor" responses to demonstrate the range of student responses with the participants and model the scoring process by "thinking aloud". Then, participants working in small groups scored five more anchor

responses independently and discussed the scores they assigned. In discussions, groups were asked to focus on inconsistent scores and attempt to reconcile them. Next, groups reported their scores to the full group and the full group discussed the remaining inconsistencies and reconciled them. This process was repeated twice. Afterwards, participants scored seventy-five student responses. Finally, they completed an open-ended comment sheet and exited the scoring session.

The ten external (non-NCSU) participants did not experience a training session. Instead they were supplied with study materials, background information, and directions via the mail. The mailing included study background material and the seventy-five study responses. The participants returned their completed study materials by mail.

Statistical Analysis

To prepare the materials for statistical analysis, the researcher assigned each cell of the study rubric a point value. The point values were subjected to quantitative analysis to describe student performance, test for interrater reliability, and explore the validity of participants' rubric use. Of the three statistical analyses, the last is most significant for demonstrating the utility of rubrics as tools for EBDM.

To establish that rubrics can produce valid analysis and evaluation of library products and services, this study used Cohen's kappa statistic and a "gold standard" approach. Typically used as a measurement of interrater reliability, Cohen's kappa can also be used to compare a group of raters to a "gold standard" to check for validity (Gwet, 2001, p. 202). Gwet (2001) explains that the gold standard is the "correct classification of subjects made by an experienced observer" (p. 202). When a gold standard approach is used, it is assumed that "the researcher knows the 'correct classification' that may be due to an expert judgment" (p. 223). Gwet (2001) explains, "The question that the researcher wants to answer is whether the…raters agree with the standard. Instead of evaluating the extent of agreement between raters, the researcher wants to know how truthful are the observers' ratings" (p. 223). This approach is also known as a "rater-to-standard reliability" or "rater-to-expert reliability" (Gwet, 2001, p. 223). Using Cohen's kappa, this study compared each participant's rubric scores to the gold standard set by the researcher. Then participants were ranked according to their

rater-to-standard reliability. For EBDM, only scores from the most "expert" or valid rubric users are fit for use.

A major benefit of this method for establishing valid rubric data is that Cohen's kappa statistic can be easily interpreted using an index established by Landis and Koch (1977) (see Figure 2) (p. 165). In this study, kappas were calculated for each participant on the four rubric criteria as well as the summary score assigned to student responses (see Figure 3).

Kappa Statistic	Strength of Agreement
<0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

Figure 2. Kappa Statistics and Strength of Agreement

Expert Rubric Users

Statistical analysis revealed that a subset of participants included in this study were able to achieve substantially valid results. Two internal (NCSU) librarians and three instructors formed the expert group, and their rubric data contributed to improved online instruction. The emergence of an expert participant group is a promising sign for the use of rubrics to support EBDM.

On the other hand, three internal librarians and all of the external librarians included in the study were unable to achieve an expert level of validity (see Figure 3). An analysis of the interrater reliability among the expert participants (see Figure 4) and non-expert participants (see Figure 5) demonstrates the distinctions between the two groups. Expert participants demonstrated moderate and substantial agreement across all criteria of the study rubric as well as the overall "grade" assigned to the student performance. Non-expert participants showed only fair or slight agreement across nearly all areas of the rubric. The application of t-tests to this data shows that these differences in reliability levels are statistically significant.

Average Kappa	Rank	Participant Group	Status
0.72	**1**	**NCSU Librarian**	**Expert**
0.69	**2**	**Instructor**	**Expert**
0.67	**3**	**Instructor**	**Expert**
0.66	**4**	**Instructor**	**Expert**
0.62	**5**	**NCSU Librarian**	**Expert**
0.61	6	Instructor	Non-Expert
0.59	7	Instructor	Non-Expert
0.58	8	Student	Non-Expert
0.56	9	Student	Non-Expert
0.55	10	NCSU Librarian	Non-Expert
.055	11	Student	Non-Expert
0.54	12	Student	Non-Expert
0.52	13	Student	Non-Expert
0.52	14	NCSU Librarian	Non-Expert
0.43	15	External Instruction Librarian	Non-Expert
0.32	16	External Reference Librarian	Non-Expert
0.31	17	External Instruction Librarian	Non-Expert
0.31	18	NCSU Librarian	Non-Expert
0.30	19	External Reference Librarian	Non-Expert
0.30	20	External Instruction Librarian	Non-Expert
0.27	21	External Reference Librarian	Non-Expert
0.21	22	External Instruction Librarian	Non-Expert
0.19	23	External Reference Librarian	Non-Expert
0.14	24	External Instruction Librarian	Non-Expert
0.13	25	External Reference Librarian	Non-Expert

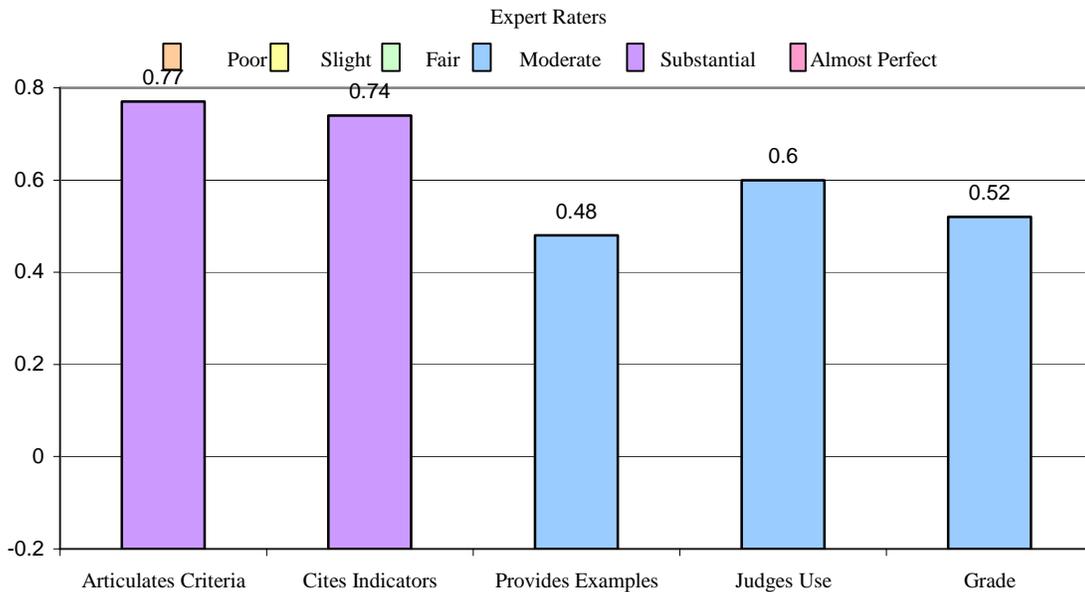*Figure 3. Rank Order of Participants by Average Kappa*

Expert Raters

| | Poor | | Slight | | Fair | | Moderate | | Substantial | | Almost Perfect |

Figure 4. Kappa Statistics for Expert Rubric Users, Standard Error =.03

Non-Expert Raters

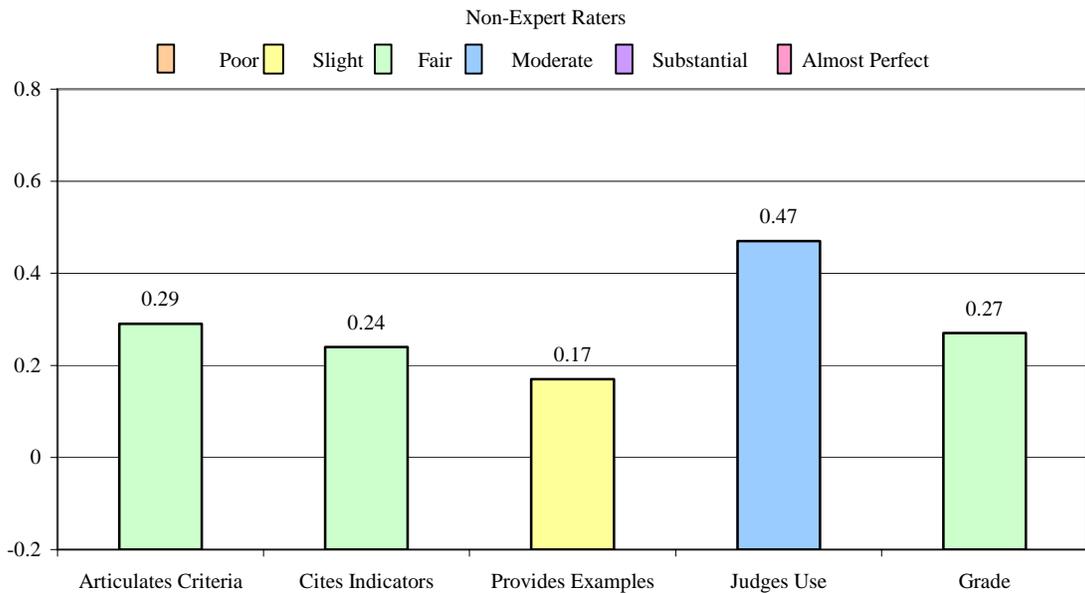| | Poor | | Slight | | Fair | | Moderate | | Substantial | | Almost Perfect |

Figure 5. Kappa Statistics for Non-Expert Rubric Users, Standard Error =.005, .006

This statistical analysis revealed two additional items worth noting. First, perhaps not surprisingly, the external librarians provided the least valid rubric results of all participants. This may be attributable to their lack of training or lack of familiarity with

Dr. Megan Oakleaf
Using Rubrics to Collect Evidence for Decision-Making: What do Librarians Need to Learn?

NCSU library contexts. The second remarkable result is the wide variation of the internal librarians' validity rankings. Although the highest validity rank was achieved by a NCSU librarian, a second NCSU librarian earned scores that were no more accurate than external librarians. This extreme variation across NCSU librarians demonstrates that training and familiarity with library culture are not enough to ensure valid rubric usage. This finding also highlights the importance of conducting basic tests to confirm the validity of data used to make decisions regarding library products and services.

Not All Rubric Users Are Experts

One of the most important findings of this study is that not all rubric users are experts. While this might seem an obvious observation, in practice many rubrics are used by those who that have not been verified as expert in the use of such tools. Often, expert ability is assumed based on an individual's educational background, experience, or position in an educational institution. However, in this study, the group of expert participants crossed divisions of all three areas. Thus, the belief that rubrics can be used reliably and validly by those who have a particular degree, a specific type of experience, or a certain position within the institution is a dangerous assumption. If rubric users are selected because of their education, experience, or position rather than demonstrated ability to provide reliable and valid scores, library processes may be evaluated inconsistently, inaccurately, or unfairly. This study supports the idea that only the rubric results of demonstrated experts should be used to make decisions that impact library products and services.

Characteristics of Expert Rubric Users

While this study offers a statistical process for identifying expert rubric users, it is also wise to consider the factors that make an "expert" an expert. Why do some rubric users achieve expert status while others do not?

Some possible answers to this question emerge both from the literature and from the comments of the participants in this study. First, the literature reports that non-expert rubric users may have diverse outlooks, perspectives, and experiences that need to be taken into account (Colton, Gao, Harris, Kolen, Martinovich-Barhite, Wang & Welch, 1997). Tired rubric users may not score students evenly over a time period. A user's

mood can affect the scores they assign. Prior knowledge can also impact rubric users' responses (Moskal, 2000).

Some rubric users might be naturally proficient scorers. Wolfe, Kao and Ranney note that the most proficient scorers tend to focus on the general features of a product or performance and "adopt values espoused by the scoring rubric" more so than less proficient scorers (Wolfe, Kao & Ranney, 1998). Throughout the scoring process, rubric users should revisit the established criteria in order to ensure that consistency is maintained (Moskal, 2000). However, less proficient rubric users tend to interrupt their observation of a product or performance to monitor how well it satisfies the rubric, rather than focusing on the product or performance and then reviewing it against the rubric (Wolfe et al., 1998).

Training can also impact the proficiency of rubric users, and this is born out by the significant differences between internal and external participants in this study. In addition to the lack of training, there may be other barriers that keep rubric users from becoming experts. In this study, participants who completed the open-ended comment sheet after scoring student responses alluded to some of these roadblocks. Their comments can be categorized into six barriers that may explain why some participants could not attain expert status.

Barrier 1: Difficulty Understanding an Outcomes-Based Approach

Libraries have a strong tradition of using input and output measures to evaluate products and services and are sometimes slow to adopt outcomes-based approaches. As a result, librarians may struggle with outcomes-based tools like rubrics. The rubric used in this study was based on outcomes included in the Information Literacy Competency Standards for Higher Education, a document well known among reference and instruction librarians. Despite the popularity of the standards and the recent push for outcomes-based assessment of information literacy skills in academic libraries, some participants of the study appeared to be unfamiliar with key concepts of outcomes-based methods. A few participants voiced concerns that an outcomes-based approach to the assessment of information literacy instruction may fail to measure what they termed student "understanding" or "ability". These participants felt that using measurable outcomes to assess student learning focused too much on specific skills—too much "science" and not

Dr. Megan Oakleaf

Using Rubrics to Collect Evidence for Decision-Making: What do Librarians Need to Learn?

enough "art". One participant wrote, "While the rubric measures the presence of concepts…it doesn't check to see if students understand [the] issues." Another participant stated, "This rubric tests skills, not…real learning." These comments indicate that librarians need to learn more about the values and principles of outcomes-based assessment before they can make meaningful progress can be made in the assessment of information literacy skills.

The need for improvement in this area has also been noted in the literature. Despite reports that librarians are increasingly asked to join other educators in justifying their programs using learning outcomes (Lichtenstein, 2000, p. 28), Iannuzzi (1999) reports, "We have yet to see widespread implementation of outcomes assessment methodologies in terms of student learning in our academic libraries" (p. 304). Despite the slow diffusion of such measures, outcomes-based approaches offer real benefits to librarians striving to use EBDM practices to advance library programs. Smith (2000) states, "It is important for libraries to understand the processes that are used to define learning outcomes, to select measures, to collaborate with other academic departments, and to use the results to improve their programs."

Barrier 2: Tension Between Analytic and Holistic Approaches

A second barrier that may have prevented some librarians in this study from attaining expert status is a lack of comfort with analytical evaluation. One participant worried that parsing user behavior into components may result in oversimplification, commenting that using the rubric "was <u>really</u> simple. But I worried that I was being too simplistic…and not rating [student work] holistically." Another participant seemed to prefer assigning an "overall" score, stating "the rubric is a good and a solid way to measure knowledge of a process but it does not allow for raters to assess the response as a whole."

These comments reveal tension between analytic and holistic approaches to evaluation. Both approaches have advantages and disadvantages. Arter and McTighe (2000) recommend holistic rubrics for simple products or performances, particularly ones with only one important criterion to assess. They also note that holistic rubrics are useful for "getting a quick snapshot of overall quality or achievement" (p. 21). Unfortunately, holistic rubrics provide only limited feedback (Mertler, 2001) and provide "no detailed

analysis of the strengths and weaknesses of a product or performance. So, they're not as useful diagnostically to help plan instruction. Nor do they provide students with detailed feedback to guide their improvement" (Arter & McTighe, 2000, p. 21).

Because analytic rubrics "divide…a product or performance into essential traits or dimensions so that they can be judged separately—one analyzes a product or performance for essential traits" (Arter & McTighe, 2000, p. 18), they allow for separate evaluations of each factor along a different descriptive scale (Moskal, 2000). This part-to-whole approach makes analytic rubrics better suited for "judging complex performances (e.g. research process) involving several significant dimensions" (Arter & McTighe, 2000, p. 22). According to Mertler (2001), analytic rubrics are preferred when a focused response to stakeholders is required. However, these advantages come at a cost. Analytical rubrics take more time to create and to use. Arter and McTighe (2000) write, "After all, you have more to discern" (p. 23).

The decision to use a holistic or analytic rubric is one that should be based on the product or performance to be assessed, the criteria to be observed, and the purpose of the assessment. Mertler states that the most important factor to consider is how the results of the assessment will be used. He writes, "If an overall, summative score is desired, a holistic scoring process would be more desirable. In contrast, if formative feedback is the goal, an analytic scoring rubric should be used" (Mertler, 2001). In the case of the study rubric, formative feedback was required, not only for students, but for librarians. Without formative feedback, librarians could not make decisions for improvements to the tutorial. In a different EBDM scenario, a holistic rubric might be more appropriate.

Barrier 3: Failure to Comprehend the Rubric

A third issue that may prevent rubric users from becoming "experts" is simple: a failure to understand some aspect of the rubric. Nearly all the participants used in this study claimed that they understood the terms used in the rubric. Still, one important comprehension problem surfaced during this study. A few external librarian participants failed to realize that rubrics are intended to judge performances on quality, not quantity or how many times a performance is observed (Callison, 2000, p. 36). One participant wrote, "I decided to use literally example<u>s</u>, indicator<u>s</u> to mean that students needed to provide more than one." Another external participant commented, "The student might

Dr. Megan Oakleaf

Using Rubrics to Collect Evidence for Decision-Making: What do Librarians Need to Learn?

cite one example…but not…enough for me to consider it exemplary." This focus on the quantity rather than quality reveals a lack of experience with rubric design that might be easily addressed through training.

Barrier 4: Disagreement with Assumptions of the Rubric

The content of any rubric is based on certain assumptions. In this study, the major assumption is that the capacity to use the criterion of authority for website evaluation is comprised of the ability to use criterion terminology, cite examples of indicators of the criterion, identify those indicators in an example website, and make a reasoned decision about the use of the example website. This assumption is based on two documents: the Information Literacy Competency Standards for Higher Education and the Objectives for Information Literacy Instruction: A Model Statement for Academic Librarians. Perhaps the biggest hurdle for some study participants was their disagreement with the standards on which the rubric was based; as a result, they could not fully adopt the assumptions of the rubric. The standard that posed the most difficulty for participants was Standard 3.2 which reads, "The information literate student articulates and applies initial criteria for evaluating both the information and its sources" (American Library Association, 2000). Several participants argued that students should not have to articulate web evaluation criteria—in this case, authority. These participants felt that assessing students' ability to use criteria terminology resulted in "looking for specific terms and missing the overall point of the responses." One participant commented that by expecting students to articulate criteria terminology, the rubric "valued students' ability to use particular words but does not measure their understanding of concepts." This comment indicates that disagreement with one part of the rubric impacted the participant's ability to adopt the values of the rubric in its entirety.

It is important to note that conflicts with the assumptions of a rubric might be avoided if participants are included in the rubric development process. For this study, such an approach was not feasible, but research confirms the value of allowing stakeholders to discuss and determine agreed upon values of student learning.

Barrier 5: Difficulties with Artifacts

Artifacts of library processes present a fifth potential barrier to participants' expert status. In this study, several participants commented on the difficulty of

interpreting artifacts of student learning because some student responses were cryptic, incomplete, vague, or incorrect. One external participant said that she found herself "giving the more cryptic answers the benefit of the doubt." Another complained that student responses were sometimes incomplete. She questioned, "If a student answer consists of a bulleted list of responses to the prompt, but no discussion or elaboration, does that fulfill the requirement?" Another lamented, "It's really hard…when students are asked to describe, explain, draw conclusions, etc. and some answer with one word." A third asked, "Should the rubric be used on 'text' that isn't in complete sentence form? How much should one use the LOBO prompt to interpret student answers?" Some participants were stymied by incorrect or partially incorrect answers. One wrote, "I suspect my own perceptions of the 'correctness' of the answers affected me." Another stated, "It killed me that I couldn't take points off for incorrect information." These comments indicate that difficulties with artifacts can affect participants' ability of to produce consistent and accurate results.

Barrier 6: Difficulties Understanding Library Context and Culture

In this study, external librarian participants achieved the lowest levels of validity, a result that could be due to lack of training or a lack of familiarity with NCSU library conventions and culture. Either way, it appears the typical model for exporting a tool that works at one library to another library, via professional listservs or journals, may not produce valid data for EBDM. This is an important finding with implications for librarians' professional practice. If tools are to be shared among libraries, it appears that local training and adaptation are necessary, and the results should be statistically analyzed to confirm data quality.

The Need for Training

In the hands of experts, the rubrics employed in this study yielded detailed, yet comprehensible data that facilitated instructional decision-making. Yet this study also reveals that non-expert rubric users may not be able to produce valid evaluations of library products and services without additional local training.

To overcome barriers to successful rubric usage, training should teach librarians the value and principles of outcomes-based approaches to library analysis and evaluation. Training should incorporate the theories that underlie rubrics and the advantages and

disadvantages of various rubric models, such as analytic and holistic rubrics. In addition, training should cover the structural issues that can limit the reliability, validity, and overall usefulness of rubrics: some rubrics are not well written (Popham, 2003, p. 95); some use wording that is too general or too specific (Tierney & Simon, 2004); some are too long (Popham, 2003, p. 98); some include inconsistencies (Tierney & Simon, 2004); and some emphasize quantity rather than quality (Callison, 2000, p. 36). Furthermore, training should address methods for eliminating disagreement with assumptions of a rubric. Finally, because rubrics may be used to analyze and evaluate library products and services that generate "messy" data, training should review the difficulties librarians are likely to encounter in the data and methods for handling them.

The Value of Rubrics

If so much training is required to help librarians use rubrics to produce reliable and valid results, skeptics might ask if the benefits are worth the time and energy training requires. Certainly, there are costs associated with training, but the advantages of rubrics outweigh the costs (Prus & Johnson, 1994, p. 25). The participants of this study confirmed the value of rubrics—nearly all participants stated that they could envision using rubrics to improve library instructional services. Such feedback attests to the merit of rubrics as tools for effective EBDM practice. In short, librarians eager to use EBDM strategies should invest in rubric training. To do otherwise is to discard a viable approach to EBDM that offers librarians a wealth of benefits.

References

American Library Association. (2000). *Information Literacy Competency Standards for Higher Education*. Retrieved April 22, 2005, from <http://www.ala.org/ala/acrl/acrlstandards/informationliteracycompetency.htm>.

Arter, J. & McTighe, J. (2000). *Scoring Rubrics in the Classroom: Using Performance Criteria for Assessing and Improving Student Performance*. Thousand Oaks, California: Corwin Press.

Bernier, R. (2004). Making yourself indispensible by helping teachers create rubrics. *CSLA Journal, 27* (2).

Bresciani, M. J., Zelna, C. L., & Anderson, J. A. (2004). *Assessing Student Learning and Development: A Handbook for Practitioners*. National Association of Student Personnel Administrators.

Callison, D. (2000). Rubrics. *School Library Media Activities Monthly, 17* (2).

Colton, D. A., Gao, X., Harris, D. J., Kolen, M. J., Martinovich-Barhite, D., Wang, T. & Welch, C. J. (1997). *Reliability Issues with Performance Assessments: A Collection of Papers*. ACT Research Report Series 97-3.

Gwet, K. (2001). *Handbook of Inter-Rater Reliability: How to Estimate the Level of Agreement between Two or Multiple Raters*. Gaithersburg, Maryland: STATAXIS.

Hafner, J. C. (2003). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education, 25* (12).

Iannuzzi, P. (1999). We are teaching, but are they learning: Accountability, productivity, and assessment. *Journal of Academic Librarianship, 25* (4).

Landis, J. R. & Koch, G. G. (1977). The measure of observer agreement for categorical data. *Biometrics, 33*.

Lichtenstein, A. A. (2000). Informed instruction: Learning theory and information literacy. *Journal of Educational Media and Library Sciences, 38* (1).

Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research and Evaluation, 7* (25).

Moskal, B. M. (2000). Scoring rubrics: What, when, and how? *Practical Assessment, Research, and Evaluation, 7* (3).

Nitko, A. J. (1996). *Educational Assessment of Students*. Englewood Cliffs, New Jersey: Prentice Hall.

Popham, J. W. (2003). *Test Better, Teach Better: The Instructional Role of Assessment*. Alexandria, Virginia: Association for Supervision and Curriculum Development.

Prus, J. & Johnson, R. (1994). A critical review of student assessment options. *New Directions for Community Colleges, 88*.

Smith, K. R. (2000). *New Roles and Responsibilities for the University Library: Advancing Student Learning through Outcomes Assessment.* Association of Research Libraries.

Stevens, D. D. & Levi, A. (2005). *Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback, and Promote Student Learning.* Sterling, Virginia: Stylus.

Tierney, R. & Simon, M. (2004). What's still wrong with rubrics: Focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research, and Evaluation, 9* (2).

Wiggins, G. (1996). Creating tests worth taking. In R. E. Blum and J. A. Arter (Eds.), *A Handbook for Student Performance in an Era of Restructuring*. Alexandria, Virginia: Association for Supervision and Curriculum Development.

Wolfe, E. W., Kao, C. & Ranney, M. (1998). Cognitive differences in proficient and nonproficient essay scorers. *Written Communication, 15* (4).