

**ASSESSING INFORMATION LITERACY SKILLS:  
A RUBRIC APPROACH**

Megan J. Oakleaf

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Information and Library Science.

Chapel Hill  
2006

Approved by  
Advisor: Helen Tibbo  
Reader: Marilee Bresciani  
Reader: Evelyn Daniel  
Reader: Barbara Moran  
Reader: Paul Solomon

© 2006  
Megan J. Oakleaf  
ALL RIGHTS RESERVED

## **ABSTRACT**

Megan J. Oakleaf

Assessing Information Literacy Skills: A Rubric Approach

(Under the direction of Dr. Helen Tibbo)

Academic librarians should explore new approaches to the assessment of information literacy skills. Satisfaction surveys and input/output measures do not provide librarians with adequate information about what students know and can do. Standardized multiple-choice tests and large-scale performance assessments also fail to provide the data librarians need to improve instruction locally. Librarians, facing accountability issues and possessing the desire to improve student learning, require a new approach to library instruction assessment.

This study investigated the viability of a rubric approach to information literacy assessment and examined an analytic information literacy rubric designed to assess students' ability to evaluate website authority. The study addressed these questions: (1) To what degree can different groups of raters provide consistent scoring of student learning artifacts using a rubric? (2) To what degree can raters provide scores consistent with those assigned by the researcher? (3) To what degree can students use authority as a criterion to evaluate websites?

This study revealed that multiple raters can use rubrics to produce consistent scoring of information literacy artifacts of student learning; however, different groups of raters in this study arrived at varying levels of agreement. For example, ENG 101 instructors produced significantly higher reliabilities than NCSU librarians and ENG 101 students, and NCSU librarians produced remarkably higher levels of agreement than external instruction and reference librarians.

In addition to providing important findings regarding the five original rater groups, this study documented the emergence of an “expert” rater group, identified through kappa statistics and a “gold standard” approach to the examination of validity. These raters not only approximated the researcher’s scores, they also achieved higher levels of agreement than any of the five original groups. This study suggests that librarians may require substantial training to overcome barriers blocking expert rater status.

Finally, this study found that most students can cite specific indicators of authority when evaluating a website. Nearly all students can locate and identify these authority indicators in a website. However, many students have difficulty choosing an appropriate website for a specific assignment and providing a rationale for their choice.

## **ACKNOWLEDGEMENTS**

Without the contributions of my colleagues, the support of my friends, and the love of my family, this dissertation would not exist. While only one name is included in the byline of this work, a number of people deserve credit for the contributions it makes to the literature of academic libraries and higher education assessment.

Among my colleagues, I must credit the LOBO tutorial team, especially Josh Boyer, who asks the right questions at the right times, and Cindy Levine, who supported my early work as an instruction librarian and shared her wisdom so that I might not run too far amuck. I wish also to thank Steve McCann for participating in the pilot test, Katherine Dexter Willis for slaying the EndNote dragon, and Herman Berkhoff for his friendship and a million favors, both large and small. For his statistical support, I thank Paul Biemer of the Odum Institute. Two members of my dissertation committee were also instrumental in this research, Dr. Evelyn Daniel, whose office door is always open, and Dr. Marilee Bresciani, who has been in all ways a friend, a mentor, and an inspiration—in the practice of assessment and in life.

Of my friends and family, I wish to thank Amy VanScoy and Kelly Seiber. Amy's contributions to this dissertation range from the cerebral work of editing drafts and double-checking data to the work only a friend knows to do...like promising a mall trip as a reward for a task accomplished or supplying a steady

stream of chocolate or caffeine as the situation merits. My gratitude also goes to Kelly, for his love, unfailing optimism, and belief in my success.

Finally, I thank my mother, Patti Owen, for her countless contributions to my life and work that have made this research possible. I thank her for reviewing every word of this dissertation...and all her help with the papers and projects leading up to it since kindergarten. I thank her for setting an example to follow as an educator and for all the lessons I have learned about librarianship at her side. I thank her for being a true friend, for knowing me better than I know myself, and for using that knowledge to push and praise me into being a better teacher, a better librarian, and—most importantly—a better person. Perhaps most of all, I thank her for reading books as a child, drawing from them a vision of how families should be, and loving me enough to make my life a fairy tale.

## TABLE OF CONTENTS

	Page
LIST OF FIGURES .....	xiv
Chapter	
1 INTRODUCTION .....	1
The Problem .....	1
The Local Problem.....	3
Purpose .....	5
Research Questions .....	6
Definition of Terms.....	7
Assumptions, Delimitations, and Limitations.....	10
Significance .....	11
2 LIBRARY INSTRUCTION IN HIGHER EDUCATION .....	13
The Role of Library Instruction in Higher Education.....	13
The Content of Library Instruction.....	15
Information Literacy Standards .....	16
Methods of Library Instruction .....	20
Instruction at the Reference Desk.....	21
Course Integrated Instruction.....	24

	For-Credit Classes .....	27
	Tutorials .....	28
3	ASSESSMENT OF LIBRARY INSTRUCTION .....	32
	Definitions of Assessment .....	32
	Purposes of Assessment .....	35
	Responding to Calls for Accountability .....	36
	Participating in Accreditation Processes .....	37
	Improving Program Performance .....	44
	Improving Program Structures .....	46
	Improving Student Learning and Teaching Skills .....	48
	Closing the Loop .....	51
	Selecting a Method of Assessment .....	52
	Utility and Relevance .....	53
	Stakeholder Needs .....	55
	Measurability and Cost .....	56
	Best Practices .....	59
	Assessment of Library Instruction: The Past .....	63
	Inputs and Outputs .....	65
	Satisfaction Surveys .....	69
	Anecdotal Reports .....	79
	The Call for Change in Assessment .....	80
	Increased Attention in Library Literature .....	83
	Current Methods of Assessing Library Instruction .....	85



	Focus on Learning Outcomes.....	86
	Outcomes-Based Education .....	90
	Outcomes-Based Assessment.....	92
	Benefits of Outcomes-Based Assessment.....	94
	Outcomes-Based Surveys and Standardized Tests ....	96
	Theoretical Background .....	98
	Benefits .....	103
	Limitations.....	105
	Multiple-Choice Items .....	117
	Standards-Based Tests.....	119
	Surveys and Tests of Information Literacy Skills .....	122
	Outcomes-Based Performance Assessments .....	134
	Constructed-Response Items.....	139
	Theoretical Background: Education .....	145
	Theoretical Background: Motivation.....	153
	Theoretical Background: Assessment.....	160
	Benefits .....	165
	Limitations.....	169
	Large-Scale Accountability Testing and Classroom Assessment .....	171
	Performance Assessments of Information Literacy Skills .....	175
4	RUBRICS AND INFORMATION LITERACY ASSESSMENT.	183
	Definitions of Rubrics.....	183

	Components of a Full Model Rubric.....	186
	General and Task-Specific Rubrics .....	189
	Holistic and Analytic Rubrics .....	190
	Benefits of Rubric Assessment.....	195
	Limitations of Rubric Assessment.....	207
	Rubric Development .....	212
	Documentation of Rubric Assessment Results .....	224
	Reporting of Rubric Assessment Results .....	229
	Rubrics of Information Literacy Skills.....	230
5	METHODOLOGY .....	265
	Study Participants.....	265
	Providers of Artifacts of Student Learning.....	266
	Raters .....	267
	The Rubric .....	271
	The Pilot Test.....	273
	Study Procedure .....	274
	Preparation of Study Materials .....	275
	The Internal Rater Experience .....	279
	The External Rater Experience .....	282
	Preparation for Statistical Analysis.....	283
	Statistical Analysis of Reliability.....	284
	Interrater Reliability .....	285
	Consensus Estimates .....	291

	Analysis of Validity .....	297
	Content-Related Validity .....	298
	Construct-Related Validity.....	300
	A “Gold Standard” Test for Validity .....	302
	Funding.....	303
6	RESULTS.....	304
	Reliability of Rubric Assessment.....	304
	Reliability Within Rater Groups .....	306
	Significant Differences Among Rater Groups....	315
	Reliability Differences Throughout the Scoring Process .....	320
	Expert Raters .....	320
	Validity of Rubric Assessment .....	325
	Student Use of Authority as a Criterion for Website Evaluation .....	327
	“Articulates Criteria” .....	328
	“Cites Indicators of Criteria” .....	329
	“Links Indicators to Examples from Source” .....	331
	“Judges Whether or Not to Use Source” .....	331
	Summary of Results .....	333
7	DISCUSSION .....	341
	Purpose of the Study .....	341

Research Questions and Answers.....	342
Consistent Scoring of Artifacts of Student Learning.....	343
Agreement Between Raters and the Researcher .....	348
Students' Use of Authority to Evaluate Websites.....	350
Limitations of the Study .....	351
Implications of the Study.....	352
Not All Raters Are Experts .....	353
A Method for Identifying Expert Raters .....	354
Characteristics of Expert Raters .....	354
Barrier 1: Difficulty Understanding Outcomes-Based Assessment.....	356
Barrier 2: Tension Between Holistic and Analytic Rubrics .....	359
Barrier 3: Failure to Comprehend the Rubric ...	363
Barrier 4: Disagreement with Assumptions of the Rubric.....	365
Barrier 5: Difficulties with Artifacts of Student Learning .....	370
Barrier 6: Difficulties Understanding Campus Context and Culture.....	371
The Need for Training.....	373
The Value of Rubrics .....	377
Recommendations for Future Research .....	381
 APPENDIX A: LOBO Information Literacy Objectives and Outcomes .....	 384
APPENDIX B: Study Rubric.....	389

APPENDIX C: Sample Score Sheet .....	390
APPENDIX D: Internal Rater Training Power Point .....	391
APPENDIX E: Selected External Rater Materials .....	401
APPENDIX F: Student Version of Study Rubric.....	409
APPENDIX G: UNC Consent Form.....	410
APPENDIX H: UNC IRB Approval .....	413
APPENDIX I: NCSU Consent Form .....	414
APPENDIX J: NCSU IRB Approval.....	416
WORKS CITED .....	417

## LIST OF FIGURES

Figure	Page
3.1 Educational Paradigms .....	100
3.2 Educational Theories .....	101
3.3 Teaching for Knowledge, Not the Test .....	113
3.4 Sample SAILS Test Question.....	133
3.5 Educational Theory and Assessment.....	152
3.6 Assessment and Motivation .....	158
3.7 Thinking Required for Successful Learning.....	164
3.8 ICT Literacy Framework.....	177
4.1 D'Angelo's Information Literacy Rubric .....	231
4.2 Merz and Mark's Information Literacy Rubric #1 .....	234
4.3 Merz and Mark's Information Literacy Rubric #2 .....	235
4.4 Merz and Mark's Information Literacy Rubric #3 .....	236
4.5 Merz and Mark's Information Literacy Rubric #4 .....	237
4.6 Rockman's Information Literacy Rubric.....	239
4.7 Emmons and Martin's Information Literacy Rubric.....	242
4.8 Buchanan's Information Literacy Rubric.....	246
4.9 Frank's Information Literacy Rubric.....	249
4.10 Gauss and Kinkema's Information Literacy Rubric.....	250
4.11 Kivel's Information Literacy Rubric .....	253
4.12 Knight's Information Literacy Rubric.....	255
4.13 Kobritz's Information Literacy Rubric.....	257

4.14	Bay Area Community Colleges' Information Literacy Rubric #1 .....	260
4.15	Bay Area Community Colleges' Information Literacy Rubric #2 .....	262
5.1	LOBO Tutorial Before Pilot Testing .....	275
5.2	LOBO Tutorial After Pilot Testing .....	276
5.3	Kappa Statistics and Strength of Agreement.....	295
5.4	Process to Ensure Content-Related Validity.....	299
6.1	Total Score to Letter Grade Conversion .....	305
6.2	Kappa Statistics and Strength of Agreement Color Codes .....	306
6.3	Kappa Statistics for NCSU Librarians .....	307
6.4	Kappa Statistics for ENG 101 Instructors.....	308
6.5	Kappa Statistics for ENG 101 Students .....	309
6.6	Kappa Statistics for External Instruction Librarians.....	311
6.7	Kappa Statistics for External Reference Librarians.....	311
6.8	Kappa Statistics for Internal Raters.....	312
6.9	Kappa Statistics for External Raters .....	313
6.10	Kappa Statistics for All Raters.....	314
6.11	Rank Order of Raters .....	321
6.12	Kappa Statistics for Expert Raters .....	322
6.13	Kappa Statistics for Non-Expert Raters.....	323
6.14	Validity Rank and Minutes Spent Scoring .....	327
6.15	Distribution of Students' Scores for "Articulates Criteria.....	329
6.16	Distribution of Students' Scores for "Cites Indicators of Criteria.....	330
6.17	Distribution of Students' Scores for "Links Indicators to Examples from Source.....	332

6.18	Distribution of Students' Scores for "Judges Whether or Not to Use the Source .....	333
------	--	-----



## **CHAPTER 1**

### **INTRODUCTION**

In the twenty-first century, educational institutions at all levels face calls for accountability. Until recently, the demands faced by other academic units on campus have passed over college and university libraries. Now, academic librarians are increasingly pressured to prove that the resources and services they provide result in improvement in student learning and development. In answer to calls for accountability, academic librarians must demonstrate their contribution to the teaching and learning goals of their institutions.

#### **The Problem**

Most educators realize that educational accountability programs are here to stay. The public—students, parents, politicians—have “serious doubts about whether educators are doing an adequate job of teaching.”<sup>1</sup> Institutions of higher education, including academic libraries, face increasing demands for proof that they educate students effectively and efficiently.<sup>2</sup> While many colleges and universities respond to calls for accountability through the assessment of courses

---

<sup>1</sup> W. James Popham, Test Better, Teach Better: The Instructional Role of Assessment (Alexandria, Virginia: Association for Supervision and Curriculum Development, 2003). 139-140.

<sup>2</sup> Robert E. Dugan and Peter Hernon, "Outcomes Assessment: Not Synonymous with Inputs and Outputs," Journal of Academic Librarianship 28.6 (2002). 380.

and programs, library instruction of information literacy skills has generally not been included in these campuswide assessment initiatives.<sup>3</sup>

Not only do academic libraries not participate in campuswide assessments, library literature reveals that they make “few rigorous efforts to evaluate the teaching of information literacy concepts and skills” independently from institutional initiatives.<sup>4</sup> This may be because librarians are not sufficiently trained in educational methods and assessment. Pausch and Popp write:

Newly graduated librarians know a great deal about computers, technology, and databases, but many know nothing about teaching patrons to use those same tools. Instruction librarians, without proper preparation, find the continuation or implementation of instructional programs to be difficult and, often, the work is done badly or not at all. Librarians already active in library instruction need continuing education in pedagogy and assessment techniques.<sup>5</sup>

Because many librarians have not received sufficient training, they need instructional assistance and examples to follow. Librarians often look to their library colleagues for guidance, but rarely tap their counterparts in education.

Lichtenstein laments:

Too often, librarians approach the design of information literacy programs without paying attention to the decades of successful work that has been accomplished by educational psychologists in understanding how people learn.... It is as though librarians think they must discover all over again the basics of learning theory that colleges of education have, for years, been teaching prospective educators. This is an inefficient approach, and it makes little sense. Instead we can build on what is already known about

---

<sup>3</sup> Lois M. Pausch and Mary Pagliero Popp, "Assessment of Information Literacy: Lessons from the Higher Education Assessment Movement," ACRL National Conference (Nashville, Tennessee: Association of College and Research Libraries, 1997).

<sup>4</sup> Pausch and Popp, "Assessment of Information Literacy."

<sup>5</sup> Pausch and Popp, "Assessment of Information Literacy."

how people learn and quickly move to apply those concepts to our information literacy efforts.<sup>6</sup>

Although Lichtenstein writes about learning theory in general, the same applies to educational assessment theory. Librarians need not discover new approaches to assessment; instead, they can access the knowledge of educational experts who have experience with a wide variety of assessment approaches.

Librarians can also learn from their counterparts in other areas of higher education. Pausch and Popp note that, as a result of accreditation processes and external calls for accountability, “much has been done in research and practice in higher education evaluation on which academic libraries can base assessment programs.”<sup>7</sup>

### **The Local Problem**

This study developed not only from the generalized need of academic librarians to assess the student learning that results from information literacy instruction, but also from a specific instance of that need. At North Carolina State University (NCSU), students are required to complete General Education Requirements (GERs) in order to graduate. According to the NCSU model for general education, students select courses from predetermined category lists. Only one course is a requirement for all students at NCSU. That course is English 101, a first-year writing course. In fact, ninety-seven percent of NCSU

---

<sup>6</sup> Art A. Lichtenstein, "Informed Instruction: Learning Theory and Information Literacy," Journal of Educational Media and Library Sciences 38.1 (2000). 22, 25-26.

<sup>7</sup> Pausch and Popp, "Assessment of Information Literacy."

students took English 101 during the 2004-2005 school year. The remaining three percent “placed out” of the course based on college admissions test scores.

Because English 101 is a GER course in the Writing, Speaking, and Information Literacy category, instructors are required to teach and assess specified learning outcomes. One of these outcomes states that students must “demonstrate critical and evaluative thinking skills in locating, analyzing, synthesizing, and using information in writing or speaking activities.”<sup>8</sup> One way in which English 101 addresses this outcome is a mandatory requirement that all English 101 students complete an online information literacy tutorial called Library Online Basic Orientation (LOBO). To fulfill this requirement, English 101 instructors integrate modules of the LOBO tutorial throughout the course. As students progress through the tutorial, they are prompted to answer open-ended questions that reinforce or extend concepts taught in the tutorial.

Given the increasing emphasis on information literacy expressed by regional accreditation bodies, librarians at NCSU anticipate that the need to assess information literacy instruction could spread quickly across the university. Past efforts to assess library instruction include experimentation with satisfaction surveys and multiple-choice exercises modeled after assessment tools at other university libraries. Recently, NCSU librarians participated in the Standardized Assessment of Information Literacy Skills (SAILS) project, but they felt that the SAILS test questions and the ways in which student scores were documented did

---

<sup>8</sup> NC State Academic Programs GER - Writing, Speaking, and Information Literacy Rationale, June 5 2005 <[http://www.ncsu.edu/provost/academic\\_programs/ger/wrtspk/rat.htm](http://www.ncsu.edu/provost/academic_programs/ger/wrtspk/rat.htm)>.

not fit their needs. NCSU librarians also participated in a pre-release pilot test of the Educational Testing Service's standardized Information and Communication Technologies (ICT) test, but rejected the ICT test based on initial perceptions that it did not provide the detailed and contextualized results needed to make improvements to information literacy instruction locally at NCSU. None of the assessment options used in the past were well suited to the needs of NCSU librarians, especially to the immediate need of English 101 instructors to assess English 101 as a GER course. Because no existing assessment tools fit the needs of NCSU librarians and English 101 instructors, a new approach to information literacy assessment is required.

### **Purpose**

The purpose of this study was to investigate the viability of a rubric approach to information literacy assessment. The study documented the use of an information literacy rubric designed to assess student information literacy skills. Rather than attempting to assess student information literacy skills holistically, the analytic rubric used in the study examined students' ability in one information literacy skill area—the ability to evaluate the authority of a website. Students exhibited this measurable behavior by responding to open-ended questions in an online tutorial. Librarians, instructors, and students served as “raters” of students' behavior by using the rubric to score their answers.

This study accomplished three major objectives. First, it analyzed the consistency with which the rubric scores are assigned by raters in five groups:

NCSU librarians, NCSU English 101 instructors, NCSU English 101 students, instruction librarians from other Association of Research Libraries (ARL) institutions, and reference librarians with some instruction responsibilities from other ARL institutions. Second, the study examined the validity of the rubric by comparing the scores given by the raters to the scores given by the researcher. Finally, the study assessed the achievement of information literacy student learning outcomes using a rubric.

### **Research Questions**

This study addressed the following questions:

1. To what degree can different groups of raters provide consistent scoring of artifacts of student learning using a rubric?
  - a. Can raters provide scores that are consistent with others in their rater group?
  - b. Can raters provide scores that are consistent across groups?
2. To what degree can raters provide scores that are consistent with scores assigned by the researcher?
3. To what degree are students able to use authority as a criterion to evaluate a website?
  - a. Are students able to use precise criteria terminology to address the authority of a website?
  - b. Are students able to cite specific indicators of authority?

- c. Are students able to cite examples of authority indicators in the websites they evaluate?
- d. Are students able to decide whether or not a site is appropriate for use and provide a rationale for their decision?

### **Definition of Terms**

For the purpose of this study, the following terms are defined below:

Assessment, according to Angelo:

Is an ongoing process aimed at understanding and improving student learning. It involves making our expectations explicit and public; setting appropriate criteria and high standards for learning quality; systematically gathering, analyzing, and interpreting evidence to determine how well performance matches those expectations and standards; and using the resulting information to document, explain, and improve performance.<sup>9</sup>

Dugan and Hernon focus the definition of assessment to the assessment of library programs. They write:

Assessment measures changes in library users as a result of their contact with an academic library's programs, resources, and services, such as student known content, developed skills and abilities, and acquired attitudes and values. Therefore, assessment is comprised of statements about what students will know/think/be able to do as a result of their contact with library programs, not statements about what the library should/could do to bring about desired outcomes.<sup>10</sup>

Information literacy is "the set of skills needed to find, retrieve, analyze, and use information."<sup>11</sup> Information literacy instruction includes "conceptual issues related

---

<sup>9</sup> Thomas A. Angelo, "Reassessing (and Defining) Assessment," AAHE Bulletin 48.3 (1995).

<sup>10</sup> Dugan and Hernon, "Outcomes Assessment." 378.

<sup>11</sup> Association of College and Research Libraries, Introduction to Information Literacy, 2003, May 15 2005  
<<http://www.ala.org/ala/acrl/acrlissues/acrlinfo/infolit/infolitoverview/introtoinfolit/introinfolit.htm>>.

to the very generation of information, the dynamics of its organization and processing and the implications of those processes for access, retrieval, and use.”<sup>12</sup> Information literacy is also “more than a framework of knowledge and a set of skills, it is an attitude that reflects an interest in seeking solutions to information problems, recognition of the importance of acquiring information skills, information confidence rather than information anxiety, and a sense of satisfaction that comes from research competence.”<sup>13</sup> Two goals of information literacy are to teach students how to learn and how to become independent learners.<sup>14</sup> According to the American Library Association Presidential Committee on Information Literacy,

Ultimately, information literate people are those who have learned how to learn. They know how to learn because they know how knowledge is organized, how to find information, and how to use information in such a way that others can learn from them. They are people prepared for lifelong learning, because they can always find information needed for any task or decision at hand.<sup>15</sup>

Outcomes “depict cognitive abilities, as well as affective dimensions that you desire your program to instill or enhance.”<sup>16</sup> Thus, outcomes are not about “what

---

<sup>12</sup> Edward K. Owusu-Ansah, "Information Literacy and Higher Education: Placing the Academic Library in the Center of a Comprehensive Solution," Journal of Academic Librarianship 30.1 (2004). 5.

<sup>13</sup> Ruth V. Small, Nasriah Zakaria and Houria El-Figuigui, "Motivational Aspects of Information Literacy Skills Instruction in Community College Libraries," College and Research Libraries 65.2 (2004). 97.

<sup>14</sup> Gary B. Thompson, "Information Literacy Accreditation Mandates: What They Mean for Faculty and Librarians," Library Trends 51.2 (2002). 231.

<sup>15</sup> American Library Association, Presidential Committee on Information Literacy: Final Report, 1989, May 15 2005 <<http://www.ala.org/ala/acrl/acrlpubs/whitepapers/presidential.htm>>.

<sup>16</sup> Marilee J. Bresciani, Carrie L. Zelna and James A. Anderson, Assessing Student Learning and Development: A Handbook for Practitioners (National Association of Student Personnel Administrators, 2004). 11.



you are going to do to the student, but rather what you want the student to know or do as a result of an initiative, course or activity.”<sup>17</sup> To be measurable, they usually include active verbs that can be identified or observed by assessors.<sup>18</sup>

Dugan and Hernon define outcomes in a library context as: “observed, reported, or otherwise quantified changes in attitudes or skills of students on an individual basis because of contact with library services, programs, or instruction.”<sup>19</sup>

Rubrics are “descriptive scoring schemes” created by educators to guide analysis of student work.<sup>20</sup> They are usually employed when educators must judge the quality of performances or constructed-response items<sup>21</sup> and can be used across a broad range of subjects.<sup>22</sup> Haffner writes, “In the educational literature and among the teaching and learning practitioners, the word ‘rubric’ is understood generally to connote a simple assessment tool that describes levels of performance on a particular task and is used to assess outcomes in a variety of performance-based contexts from kindergarten through college.”<sup>23</sup> Assessment

---

<sup>17</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 11.

<sup>18</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 11.

<sup>19</sup> Dugan and Hernon, "Outcomes Assessment." 379.

<sup>20</sup> Barbara M. Moskal, "Scoring Rubrics: What, When, and How?," Practical Assessment, Research, and Evaluation 7.3 (2000).

<sup>21</sup> Popham, Test Better, Teach Better. 95.

<sup>22</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>23</sup> John C. Hafner, "Quantitative Analysis of the Rubric as an Assessment Tool: An Empirical Study of Student Peer-Group Rating," International Journal of Science Education 25.12 (2003). 1509.

experts proclaim rubrics as “one of the most basic tools in the performance assessor’s kit.”<sup>24</sup> Wiggins defines this core assessment tool as:

A set of scoring guidelines for evaluating students’ work. Rubrics answer the following questions: By what criteria should performance be judged? Where should we look and what should we look for to judge performance success? What does the range in the quality of performance look like? How do we determine validity, reliability, and fairly what score should be given and what that score means? How should the different levels of quality be described and distinguished from one another?<sup>25</sup>

### **Assumptions, Delimitations, and Limitations**

The underlying assumption of this study is that information literacy skills can be expressed as student learning outcomes, described in sufficient detail to be captured in artifacts of student learning, and recognized and scored by raters. It is also assumed that the ability to use the criterion of authority to evaluate of a website 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.

A delimitation of this study is that the participants include English 101 students, English 101 instructors, and librarians at North Carolina State University and a selected group of librarians from other universities. Students and instructors in other classes at North Carolina State University and at other universities have not been included. Librarians at other universities, with the exception of the selected group mentioned above, have not been included.

---

<sup>24</sup> Grant Wiggins, Educative Assessment: Designing Assessments to Inform and Improve Student Performance (San Francisco: Jossey-Bass, 1998). 153.

<sup>25</sup> Wiggins, Educative Assessment. 154.

Therefore, the findings are not necessarily generalizable to students and instructors in other areas at North Carolina State University, nor are they necessarily generalizable to librarians at other universities.

A limitation of this study is that students, having received instruction about how to use the criterion of authority to evaluate a website, may exhibit better than normal behavior due to temporary heightened awareness.

### **Significance**

Assessing student learning is a rapidly growing focus of institutions of higher education. If libraries intend to remain relevant on campus, they too must demonstrate their contributions to the mission of the institution by becoming involved in assessment.<sup>26</sup> Assessment provides a way to argue for the importance of libraries in colleges and universities and position reference and instruction librarians as full partners in the institution.<sup>27</sup> In order to leverage the full power of assessment, however, academic librarians need to find assessment tools that match their needs. Existing library instruction assessment tools do not meet the current needs of most academic librarians. For example, satisfaction surveys do not measure most student learning outcomes. In addition, selected-response standardized tests do not adequately measure the achievement of higher-level thinking skills. Furthermore, the performance-based ICT test

---

<sup>26</sup> Lorrie A. Knight, "The Role of Assessment in Library User Education," Reference Services Review 30.1 (2002).

<sup>27</sup> James K. Elmborg, "Teaching at the Desk: Toward a Reference Pedagogy," Portal: Libraries and the Academy 2.3 (2002). 463.

emphasizes computer literacy more than information literacy and lacks the campus context so critical to effective assessment. All of these options fail to deliver results documentation that can be used to improve information literacy instruction efforts. As a result, academic librarians need a new approach to information literacy instruction assessment.

Rubric assessment may be a good match for library instruction assessment needs. It is well suited to measure student learning outcomes, especially those that focus on higher-level thinking skills. Rubrics can be developed that are both general enough to be shared by multiple groups but yet analytic enough to apply to specific instructional activities. A few academic librarians have experimented with rubrics; however, none of them investigated the reliability or validity of a rubric approach to information literacy assessment. Such an investigation is warranted.

The significance of this study lies in its scope and approach. Thus far, there are no studies in the library literature in which a large number of raters have used rubrics to score a large number of artifacts of student learning. There are no studies in the library literature that analyze the interrater reliability of multiple raters among and/or between groups. Also, there are no studies that examine the validity of a rubric used to assess information literacy skills. Finally, on a local level, this study contributes significantly to a growing understanding of the ability of NCSU students to use authority as a criterion for evaluating websites.

## **CHAPTER 2**

### **LIBRARY INSTRUCTION IN HIGHER EDUCATION**

This chapter outlines the role of library instruction in institutions of higher education. It also describes the content of library instruction—information literacy skills. Finally, this chapter details the major methods through which information literacy skills are taught. These methods include the reference desk, course-integrated instruction, for-credit classes, and tutorials.

#### **The Role of Library Instruction in Higher Education**

In higher education, the role of library instruction is to create life-long learners, and the path to that goal is through information literacy skills.<sup>28</sup> Two goals of information literacy instruction are to teach students how to learn and how to become independent learners.<sup>29</sup> These two goals represent the “core mission of all education.”<sup>30</sup> In recent years, many educators have realized that producing information literate students is a goal that they “can neither ignore nor openly refuse a need to achieve.”<sup>31</sup> Educators across campus recognize that information literacy skills are fundamental to the development of competent

---

<sup>28</sup> Sue Samson, "What and When Do They Know? Web-Based Assessment," Reference Services Review 28.4 (2000).

<sup>29</sup> Thompson, "Information Literacy Accreditation Mandates." 231.

<sup>30</sup> Owusu-Ansah, "Information Literacy and Higher Education." 12.

<sup>31</sup> Owusu-Ansah, "Information Literacy and Higher Education." 4.

students and graduates.<sup>32</sup> As a result, information literacy instruction has become “the critical campuswide issue for the 21<sup>st</sup> century, of keen importance to all educational stakeholders, including faculty, librarians, and administrators.”<sup>33</sup>

Within higher education, academic libraries are “uniquely positioned to provide the resources that teach students how to learn within the university curriculum and then carry those applications forward in a highly competitive, information-rich society.”<sup>34</sup> Academic libraries form the cornerstone of information literacy instruction because they are the segment of the educational institution that is already focused on information problems. Librarians are familiar with information issues, are committed to teaching students to be information literate, and have the expertise to do so.<sup>35</sup>

In fact, librarians have always sought to empower students to learn independently and, despite the opinions of others or themselves, they have always been teachers.<sup>36</sup> In the past, librarians were not always expected to teach, and although job postings sometimes listed teaching experience as desirable, it was rarely included as a required qualification.<sup>37</sup> Now, librarians “increasingly recognize that their primary role is to participate in the teaching

---

<sup>32</sup> Owusu-Ansah, "Information Literacy and Higher Education." 12.

<sup>33</sup> Ilene F. Rockman, "Integrating Information Literacy into the Learning Outcomes of Academic Disciplines," College and Research Libraries News 64.9 (2003). 612.

<sup>34</sup> Samson, "What and When Do They Know? Web-Based Assessment."

<sup>35</sup> Owusu-Ansah, "Information Literacy and Higher Education." 4.

<sup>36</sup> Owusu-Ansah, "Information Literacy and Higher Education." 4.

<sup>37</sup> Lichtenstein, "Informed Instruction." 22.

missions of their institutions”<sup>38</sup> and are “racing to repackage themselves as accomplished teachers.”<sup>39</sup> According to Owusu-Ansah, “Librarians...must accept formally their teaching role and engage actively in it, not sporadically, but as a generally accepted mandate of the profession and of the academic library in academe.”<sup>40</sup> Because of the need to produce information literate students, library instruction is “not a frill or a desirable extra component, but rather an intrinsic part of education today.”<sup>41</sup>

### **The Content of Library Instruction**

Librarians have offered many names for library instruction, including bibliographic instruction.<sup>42</sup> In recent years, library instruction is referred to as “information literacy” instruction.<sup>43</sup> Whatever the name, the core concerns are the access and use of information. Information literacy teaches students “the set of skills needed to find, retrieve, analyze, and use information.”<sup>44</sup> Information literacy instruction includes “conceptual issues related to the very generation of information, the dynamics of its organization and processing and the implications

---

<sup>38</sup> Elmborg, "Teaching at the Desk." 455.

<sup>39</sup> Lichtenstein, "Informed Instruction."

<sup>40</sup> Owusu-Ansah, "Information Literacy and Higher Education." 12.

<sup>41</sup> Thompson, "Information Literacy Accreditation Mandates." 227.

<sup>42</sup> Owusu-Ansah, "Information Literacy and Higher Education." 12.

<sup>43</sup> Lichtenstein, "Informed Instruction." 23.

<sup>44</sup> Association of College and Research Libraries, Introduction to Information Literacy.

of those processes for access, retrieval, and use.”<sup>45</sup> Information literacy is also “more than a framework of knowledge and a set of skills, it is an attitude that reflects an interest in seeking solutions to information problems, recognition of the importance of acquiring information skills, information confidence rather than information anxiety, and a sense of satisfaction that comes from research competence.”<sup>46</sup> According to the American Library Association Presidential Committee on Information Literacy,

Ultimately, information literate people are those who have learned how to learn. They know how to learn because they know how knowledge is organized, how to find information, and how to use information in such a way that others can learn from them. They are people prepared for lifelong learning, because they can always find information needed for any task or decision at hand.<sup>47</sup>

### **Information Literacy Standards**

In education, standards are “verbal statement[s] of goals or desired classes of outcomes. They describe what the goals of the system are.”<sup>48</sup> In academic disciplines, content standards are domain-specific and usually written by professional organizations.<sup>49</sup> By outlining the goals of instruction, standards provide practitioners with guides to effective practice and to some degree

---

<sup>45</sup> Owusu-Ansah, "Information Literacy and Higher Education." 5.

<sup>46</sup> Small, Zakaria and El-Figuigui, "Motivational Aspects of Information Literacy Skills Instruction." 97.

<sup>47</sup> American Library Association, Presidential Committee.

<sup>48</sup> Eva L. Baker and Harold F. O'Neil Jr., "Standards for Student Learning," Encyclopedia of Education, ed. James W. Guthrie, vol. 6 (New York: Macmillan Reference USA, 2003). 2315.

<sup>49</sup> Baker and O'Neil Jr., "Standards for Student Learning." 2315.



eliminate the need for practitioners to develop their own “best practices”.<sup>50</sup>

According to Miller, “Using professional standards to guide practice and to assess the impact of that practice will benefit any and all higher education programs and services designed to enhance student learning and development.”<sup>51</sup>

In academic libraries, the Information Literacy Competency Standards for Higher Education are the professional standards that guide the practice of information literacy instruction. Developed by the Association of College and Research Libraries (ACRL) in 2000 and endorsed by the American Association for Higher Education (AAHE) and the Council of Independent Colleges, these standards offer guidance to librarians who plan to deliver and assess information literacy instruction to students.<sup>52</sup> Comprised of five standards, twenty-two performance indicators, and eighty-seven outcomes that include both higher- and lower-level thinking skills<sup>53</sup>, the outcomes are intended to guide librarians in the development of information literacy skill assessment methods, both in general, and in the context of individual academic disciplines.<sup>54</sup> The assessment of information literacy outcomes should then “reach all students, pinpoint areas for

---

<sup>50</sup> Ted K. Miller, Using Professional Standards for Program Assessment and Development, 2002, National Association of Student Personnel Administrators, 18 Feb. 2005 <<http://www.naspa.org/membership/mem/nr/article.cfm?id=624>>.

<sup>51</sup> Miller, Using Professional Standards for Program Assessment.

<sup>52</sup> American Library Association, Information Literacy Competency Standards for Higher Education, 2000, April 22 2005 <<http://www.ala.org/ala/acrl/acrlstandards/informationliteracycompetency.htm>>.

<sup>53</sup> American Library Association, Information Literacy Competency Standards.

<sup>54</sup> American Library Association, Information Literacy Competency Standards.

further program development, and consolidate learning goals already achieved. It also should make explicit to the institution's constituencies how information literacy contributes to producing educated students and citizens."<sup>55</sup> These ACRL standards are well known on university campuses and are recognized by at least one accrediting body.<sup>56</sup>

Standards offer a path of hope for education reform and instructional improvements. Standards can be used as a "communication device to rally educators and the public."<sup>57</sup> However, standards-based approaches to education are most effective when the standards are tied programmatically, substantively, and systematically to assessment measures so that realistic priorities can be set and achieved.<sup>58</sup> The American Association of Higher Education states:

Assessment works best when the programs it seeks to improve have clear explicitly stated purposes. Assessment is a goal-oriented process. It entails comparing educational performance with educational purposes and expectations—those derived from the institution's mission, from faculty intentions in program and course design, and from knowledge of students' own goals. Where program purposes lack specificity or agreement, assessment as a process pushes a campus toward clarity about where to aim and what standards to apply; assessment also prompts attention to where and how program goals will be taught and learned. Clear, shared, implementable goals are the cornerstone for assessment that is focused and useful.<sup>59</sup>

---

<sup>55</sup> American Library Association, Information Literacy Competency Standards.

<sup>56</sup> Thompson, "Information Literacy Accreditation Mandates." 222.

<sup>57</sup> Baker and O'Neil Jr., "Standards for Student Learning." 2317-2318.

<sup>58</sup> Baker and O'Neil Jr., "Standards for Student Learning." 2317-2318.

<sup>59</sup> American Association of Higher Education, Nine Principles of Good Practice for Assessing Student Learning, 1996, 28 Feb. 2005 <<http://www.aahe.org/assessment/principles.htm>>.

Standards clarify the goals of teaching and learning and the purposes of assessment. Educators use standards to “describe the ‘what’ of education further, ...to give examples of tasks that fit particular standards, ...[to] communicate more clearly the intention of general content or skill standards, ...[or] link up to the design of assessments.”<sup>60</sup> Standards also describe the “degree of proficiency, or the ‘how much’ part of performance.”<sup>61</sup> It is important to note that in all cases, the underlying theory of standards-based education is criterion-referencing.<sup>62</sup> In other words, students’ performances on assessments should be measured against standards, not other students.

While standards-based education and assessment offer many advantages, educators have noted that standards-based approaches are not without problems. One problem is that the professional organizations that write educational standards often overload educators with the high numbers of standards that cannot all be taught nor assessed.<sup>63</sup> Popham writes:

When subject-matter specialists are asked to isolate what students need to know about their specialty, the specialists’ response will almost always be “Everything!” In other words, because the people who take part in the identification of content standards typically revere their subject fields, they often identify all the knowledge and all the skills that they wish well-taught children at a given age would possess.<sup>64</sup>

---

<sup>60</sup> Baker and O’Neil Jr., “Standards for Student Learning.” 2315.

<sup>61</sup> Baker and O’Neil Jr., “Standards for Student Learning.” 2316.

<sup>62</sup> Baker and O’Neil Jr., “Standards for Student Learning.” 2316.

<sup>63</sup> Popham, Test Better, Teach Better. 32.

<sup>64</sup> Popham, Test Better, Teach Better. 32.

Long lists of standards cause problems because all the standards cannot be covered in the instructional time educators have with students, and they cannot be satisfactorily assessed on a typical test. Tests that attempt to cover such large ground often fail to test some standards, or test them with only one item.<sup>65</sup> As a result, teachers can't use the results of such tests to make defensible instructional decisions.<sup>66</sup> Popham recommends that lengthy lists of standards be revised to include fewer, more significant standards so that each could be taught well and assessed on a standard-by-standard basis that would provide detailed data to educators and administrators.<sup>67</sup>

### **Methods of Library Instruction**

The goal of library instruction is to "impart skills and knowledge that enrich and empower students in their learning and research engagements."<sup>68</sup> To accomplish this goal, teaching librarians employ a wide variety of instructional methods. A survey of library literature reveals a number of instructional approaches. Pausch and Popp list methods including single-class instruction sessions, credit courses, handouts, and online tutorials.<sup>69</sup> Owusu-Ansah describes instructional methods such as reference transactions, library tours,

---

<sup>65</sup> Popham, Test Better, Teach Better. 32.

<sup>66</sup> Popham, Test Better, Teach Better. 33.

<sup>67</sup> Popham, Test Better, Teach Better. 34.

<sup>68</sup> Owusu-Ansah, "Information Literacy and Higher Education." 12.

<sup>69</sup> Pausch and Popp, "Assessment of Information Literacy."

drop-in workshops, and general education requirements.<sup>70</sup> Grassian and Kaplowitz report a long list of methods including signage, maps, site maps, kiosks, guided tours, self-guided tours, virtual tours, exhibits, flipcharts, blackboards, whiteboards, overhead transparencies, presentation slide shows, slides/tapes and videotapes, point-of-use guides, pathfinders, exercises, computer assisted instruction, reference questions, individual research consultations, course-integrated or standalone one-shot group sessions, formal courses, discussion boards, chat, e-mail/listservs, and web pages/sites.<sup>71</sup> Feinberg and King add a workbook approach to the long list of methods.<sup>72</sup> This wide variety of instructional approaches demonstrates the flexibility and creativity of teaching librarians. Clearly, there is no “one size fits all” model program for library instruction<sup>73</sup>, and “no single method has been established as best.”<sup>74</sup>

### ***Instruction at the Reference Desk***

Although its instructional importance is sometimes overlooked, a significant amount of library instruction takes place at the reference desk. Elmborg writes, “As a staging area from which to launch into the multi-voiced,

---

<sup>70</sup> Owusu-Ansah, "Information Literacy and Higher Education."

<sup>71</sup> Esther S. Grassian and Joan R. Kaplowitz, Information Literacy Instruction (New York: Neal-Schuman, 2001). 171-208.

<sup>72</sup> Richard Feinberg and Christine King, "Performance Evaluation in Bibliographic Instruction Workshop Courses: Assessing What Students Do as a Measure of What They Know," Reference Services Review 20.2 (1992).

<sup>73</sup> Debbie Malone and Carol Videon, First Year Student Library Instruction Programs, Clip Note #33 (Chicago: American Library Association, 2003). 12.

<sup>74</sup> Mollie D. Lawson, "Assessment of a College Freshman Course in Information Resources," Library Review 48.2 (1999).

multi-genred array of resources that can be used to create knowledge, the library has no equal. When viewed this way, the reference desk can be seen as the most dynamic teaching position in the academy."<sup>75</sup> Grassian and Kaplowitz concur, stating that:

At the reference desk (physical or virtual), in-person, by phone, mail, e-mail, fax, or the Web, users get answers to their questions, but reference interactions can be instructional as well. Users may learn about the range of information sources that might answer their questions, how to evaluate and choose among them for specific information needs, and how to use them. The reference librarian has, of necessity, turned into a teacher, as well as an advisor and guide through the information morass.<sup>76</sup>

Elmborg reinforces the significance of the reference desk as an instructional method by pointing out that "for librarians with faculty status, the teaching required for the pursuit of tenure has been broadly interpreted to include reference work."<sup>77</sup>

Because reference work is a form of teaching, librarians should strive to make the most of the instructional potential of reference interactions. Elmborg suggests that librarians consider reference transactions as "academic conferences where teaching and learning take place."<sup>78</sup> He believes that reference librarians should be "coaches and collaborators" who work with student's prior knowledge and "develop the framework of understanding already begun by the patron."<sup>79</sup> Beck and Turner point out that librarians have an

---

<sup>75</sup> Elmborg, "Teaching at the Desk." 463.

<sup>76</sup> Grassian and Kaplowitz, Information Literacy Instruction. 194.

<sup>77</sup> Elmborg, "Teaching at the Desk." 456.

<sup>78</sup> Elmborg, "Teaching at the Desk." 455.

<sup>79</sup> Elmborg, "Teaching at the Desk." 463, 457.

instructional responsibility to encourage students to apply problem-solving skills and understand how information is structured and retrieved.<sup>80</sup> They further underscore the instructional importance of thinking aloud, modeling information decision-making processes, and using analogies to connect new knowledge with material the students already know.<sup>81</sup> Beck and Turner suggest the use of physical behaviors, including moving from behind the reference desk and using handouts, to enhance the instructional potential of reference transactions.<sup>82</sup>

The reference desk offers many advantages as a method of library instruction. First, instruction provided at the reference desk is “‘authentic’ in that the student has a specific project underway and has specific questions regarding how to proceed.”<sup>83</sup> Students are also most open to learning at their time of need, the time when many students visit the reference desk.<sup>84</sup> New models of reference desk service, including email and chat reference, provide additional instructional benefits. For example, students can get specific, individual instruction that is available regardless of time and location.<sup>85</sup> On the other hand, one important limitation on the instructional value of the reference desk is the brevity of transactions.<sup>86</sup>

---

<sup>80</sup> Susan E. Beck and Nancy B. Turner, "On the Fly BI: Reaching and Teaching from the Reference Desk," Reference Librarian 72 (2001). 88, 84, 85-86.

<sup>81</sup> Beck and Turner, "On the Fly BI." 87.

<sup>82</sup> Beck and Turner, "On the Fly BI." 88-89.

<sup>83</sup> Elmborg, "Teaching at the Desk." 458.

<sup>84</sup> Beck and Turner, "On the Fly BI." 83.

<sup>85</sup> Grassian and Kaplowitz, Information Literacy Instruction. 195.

<sup>86</sup> Owusu-Ansah, "Information Literacy and Higher Education." 11.

## ***Course-Integrated Instruction***

Course-integrated library instruction is often the direct result of a library research assignment given in an academic course. Many faculty believe that students who visit the library to receive instruction do better research assignments.<sup>87</sup> As a result, “library instruction generally...functions as a component within another department’s credit class.”<sup>88</sup> Grassian and Kaplowitz explain:

Course-integrated sessions may be termed “guest lectures” by an instructor, and may take place in the physical library, a computer lab, an electronic classroom, or in the normal class meeting space. Both of these types of one-shot sessions require advance planning and preparation, including goal-setting, pre- and post-testing, timing, and careful attention to the amount of material presented.<sup>89</sup>

Grassian and Kaplowitz note that these “one-shot” instructional sessions may last from ten minutes to a half or full day.<sup>90</sup> Most of the time, the librarians meet with students in-person only one time.<sup>91</sup> Dewald also offers a description of course-integrated instruction:

Classroom-based library instruction is usually by arrangement with the teaching faculty or in-group workshops offered in the library, for the purpose of providing instruction to whole classes in skills for accomplishing course projects. Classroom-based library instruction may take place in the library, computer center, or a regular classroom, and its

---

<sup>87</sup> Thompson, "Information Literacy Accreditation Mandates." 227.

<sup>88</sup> Samson, "What and When Do They Know? Web-Based Assessment."

<sup>89</sup> Grassian and Kaplowitz, Information Literacy Instruction. 199.

<sup>90</sup> Grassian and Kaplowitz, Information Literacy Instruction. 198.

<sup>91</sup> Grassian and Kaplowitz, Information Literacy Instruction. 198.



audience is a group of students who interact with each other and the librarian as well as, perhaps, computers or library materials.<sup>92</sup>

Course-integrated instruction is very common among libraries in higher education. In a survey of library instruction of first-year students, Malone and Videon found that librarians in “schools with 1000 or fewer students taught an average of 30.5 sessions per year; schools with 1001-3000 students taught 39.3 sessions per year; schools with 3001-5000 students taught 53.4 sessions; and larger schools taught an average of 102.9 sessions.”<sup>93</sup> Overall they found that course-integrated instruction was the most frequently reported type of library instruction, conducted by 84.3% of the 129 institutions they surveyed.<sup>94</sup> Interestingly, most course-integrated library instruction emphasizes the second standard of the Information Literacy Competency Standards for Higher Education, which focuses on accessing information,<sup>95</sup> and only rarely addresses the fourth and fifth standards, which focus on ethical use of information.<sup>96</sup>

Owusu-Ansah believes that, along with for-credit library courses, course-integrated instruction is the “most viable vehicle...for delivering information literacy instruction.”<sup>97</sup> Grassian and Kaplowitz highlight a number of important benefits of this method of instruction. Course-integrated instruction reaches

---

<sup>92</sup> Nancy H. Dewald, "Transporting Good Library Instruction Practices into the Web Environment: An Analysis of Online Tutorials," Journal of Academic Librarianship 25.1 (1999).

<sup>93</sup> Malone and Videon, First Year Student Library Instruction Programs. 3.

<sup>94</sup> Malone and Videon, First Year Student Library Instruction Programs. 4.

<sup>95</sup> American Library Association, Information Literacy Competency Standards.

<sup>96</sup> Small, Zakaria and El-Figuigui, "Motivational Aspects of Information Literacy Skills Instruction." 106.

<sup>97</sup> Owusu-Ansah, "Information Literacy and Higher Education." 11.

learners when they are extrinsically motivated to learn, provides opportunities for students to have personal interactions with librarians (perhaps lessening their anxiety), and is less rigorous to prepare for than for-credit courses.<sup>98</sup>

However, there are many limitations to a course-integrated approach to library instruction. Grassian and Kaplowitz point out that students who are required to attend a course-integrated session may not be intrinsically motivated, and therefore retain fewer concepts.<sup>99</sup> While not as onerous as a for-credit class, course-integrated lessons require much librarian preparation for a low number of students.<sup>100</sup> Also, students who are required to attend more than one lesson may be subjected to repeated content.<sup>101</sup> Lawson relays a concern that “a single session with a librarian who attempts to orient an entire class in how to do library research would appear to do very little.”<sup>102</sup> Thompson identifies drawbacks to course-integrated instruction. He notes that the ability to connect with students requires faculty interest, which is not always present.<sup>103</sup> Thompson also points out that some librarians, given a rare opportunity to teach students, overload students with information.<sup>104</sup> Samson too cites the “embedded nature” of library instruction as a major challenge.

---

<sup>98</sup> Grassian and Kaplowitz, Information Literacy Instruction. 199.

<sup>99</sup> Grassian and Kaplowitz, Information Literacy Instruction. 199.

<sup>100</sup> Grassian and Kaplowitz, Information Literacy Instruction. 199.

<sup>101</sup> Grassian and Kaplowitz, Information Literacy Instruction. 199.

<sup>102</sup> Lawson, "Assessment of a College Freshman."

<sup>103</sup> Thompson, "Information Literacy Accreditation Mandates." 227.

<sup>104</sup> Thompson, "Information Literacy Accreditation Mandates." 227.

### ***For-Credit Classes***

Some library instruction methods are not embedded in another course, but rather stand alone. Formal for-credit courses are usually a series of classes with the same students over a set time period, such as an academic term.<sup>105</sup> They may be synchronous or asynchronous.<sup>106</sup> Malone and Videon report that few institutions offer for-credit classes for first-year students:

Six institutions provide a required stand-alone course and ten institutions provide an elective stand-alone course. In two of these institutions the course is required for some majors and not required for others. A number of institutions said that they provide credit courses which are open to upper level students as well as first-year students, and we did not include them in these totals.<sup>107</sup>

For-credit classes have many advantages over other instruction approaches. For example, for-credit classes can be shown to increase students' knowledge and skills.<sup>108</sup> They provide opportunities to cover material in-depth, develop rapport with students, and diagnose weaknesses (and then teach or re-teach to eliminate them).<sup>109</sup> Of course, for-credit classes also have limitations. Initially, there are high costs in both staff and time.<sup>110</sup> Updating classes requires additional commitment of staff and time.<sup>111</sup> Finally, for-credit courses that are not

---

<sup>105</sup> Grassian and Kaplowitz, Information Literacy Instruction. 200.

<sup>106</sup> Grassian and Kaplowitz, Information Literacy Instruction. 200.

<sup>107</sup> Malone and Videon, First Year Student Library Instruction Programs. 4.

<sup>108</sup> Lawson, "Assessment of a College Freshman."

<sup>109</sup> Grassian and Kaplowitz, Information Literacy Instruction. 200.

<sup>110</sup> Grassian and Kaplowitz, Information Literacy Instruction. 201.

<sup>111</sup> Grassian and Kaplowitz, Information Literacy Instruction. 201.

required may be perceived as “just-in-case” rather than “just-in-time” and therefore result in low student motivation.<sup>112</sup>

### ***Tutorials***

As a result of increasingly tight higher education budgets, librarians find themselves with fewer colleagues, resources, and hours to meet the needs of students completing library research assignments. Dewald writes, “Libraries cannot afford to hire more librarians, yet the need for information instruction is greater than ever. Academic libraries are beginning to meet these challenges by using the online environment to provide new opportunities in...instruction.”<sup>113</sup> By and large, librarians develop online tutorials to bear the weight of heavy teaching loads. Tutorials may be online versions of “point-of-use guides, or holistic guides, or a conglomeration of both.”<sup>114</sup> Some tutorials include “interaction, critical thinking, and evaluation regarding a range of information sources, as well as guides to citation style, information about plagiarism, preservation, and the ethics of information use.”<sup>115</sup> Malone and Videon report that thirty-two of the 153 institutions surveyed reported using an online tutorial as a part of their library instruction for first-year students.<sup>116</sup>

---

<sup>112</sup> Grassian and Kaplowitz, Information Literacy Instruction. 201.

<sup>113</sup> Dewald, "Transporting Good Library Instruction Practices." 12-13.

<sup>114</sup> Grassian and Kaplowitz, Information Literacy Instruction. 190.

<sup>115</sup> Grassian and Kaplowitz, Information Literacy Instruction. 190-191.

<sup>116</sup> Malone and Videon, First Year Student Library Instruction Programs. 5.

Like other methods of library instruction, tutorials have benefits and limitations. One benefit of using an online tutorial is that information reaches students on a “just-in-time” rather than “just-in-case” basis.<sup>117</sup> Tutorials operate independent of time and place, and some offer immediate feedback to learners.<sup>118</sup> Students can use only what parts they need of a tutorial, jumping from topic to topic or printing out materials to use in their research at a later time.<sup>119</sup>

Despite these advantages, tutorials have two major limitations. The first limitation is the time required for maintaining and updating the content of the tutorial.<sup>120</sup> A more complex problem is evaluating the effectiveness of a tutorial. Smith notes that, often, tutorial effectiveness is not measured. She writes, “The evaluation process is often part of the design and development cycle that falls by the wayside. Whether it’s due to time or budget constraints or lack of knowledge on how to proceed, it’s not unusual to hear of projects that are not evaluated in any fashion or that have chosen the wrong measurements altogether.”<sup>121</sup> However, Smith provides only this guidance on the evaluation of online library tutorials:

The most vital goal of an online instruction project is for learning to occur.... As you are planning evaluation procedures, build in methods to

---

<sup>117</sup> Grassian and Kaplowitz, Information Literacy Instruction. 191.

<sup>118</sup> Grassian and Kaplowitz, Information Literacy Instruction. 191.

<sup>119</sup> Grassian and Kaplowitz, Information Literacy Instruction. 191.

<sup>120</sup> Grassian and Kaplowitz, Information Literacy Instruction. 191.

<sup>121</sup> Susan Sharpless Smith, Web-Based Instruction: A Guide for Libraries (Chicago: American Library Association, 2001). 170.

measure if students have mastered the content and to what extent. Even if your institution or the class instructor doesn't require a grade, a testing system will help you assess how effective your online instruction has been. This can be a crucial factor in decision making for future online instruction projects.<sup>122</sup>

Such a recommendation is helpful, but concrete advice on how to create such an evaluation is elusive. Knight cites two studies of the effectiveness of online library instruction tutorials; one found that tutorials may be more effective than in-person instruction, but the other found no significant difference that could be attributed to the format of instruction.<sup>123</sup> Further research on the assessment of learning achieved through library tutorials is definitely warranted.

While all of these instructional methods can be assessed<sup>124</sup>, Carter and others decry the general lack of literature on the effectiveness of library instruction.<sup>125</sup> This weakness in the literature of library instruction could be due to a number of factors. Perhaps most librarians do not evaluate their instruction. Or perhaps librarians evaluate their instruction, but don't often report their findings. It seems possible that the sentiments expressed by higher education student affairs personnel might apply. Bresciani, Zelna, and Anderson write:

When we first began to get serious about assessment, we approached it with a bit of arrogance. How hard could this be?, we asked. We believed assessment was common sense, and we felt we had been doing it for years—at least the part where we asked if the program had produced what we expected upon its completion and why or why not. We even administered satisfaction questionnaires to program participants and used

---

<sup>122</sup> Smith, Web-Based Instruction.

<sup>123</sup> Knight, "The Role of Assessment in Library User Education."

<sup>124</sup> Pausch and Popp, "Assessment of Information Literacy."

<sup>125</sup> Elizabeth W. Carter, "'Doing the Best You Can with What You Have': Lessons Learned from Outcomes Assessment," Journal of Academic Librarianship 28.1 (2002). 36.

that information, along with our observations, to make improvements in the next cycle. Throughout this process of asking if the program ‘worked,’ we never took the time to document the intended outcomes of the program prior to the start of the program. Nor did we record the findings in a manner that could tie to any sort of outcomes. Furthermore, we did not write down the decisions we made to improve the program. We did not even document and celebrate program successes. No one even asked us for the information.<sup>126</sup>

---

<sup>126</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 81.

## **CHAPTER 3**

### **ASSESSMENT OF LIBRARY INSTRUCTION**

This chapter provides definitions of assessment, describes purposes for conducting assessment, and identifies methods for selecting assessment tools. It also summarizes the history of library instruction assessment, including past approaches and the lack of early literature. The latter part of this chapter describes the recent increase of literature on library instruction assessment and the theoretical underpinnings of new outcomes-based performance assessment approaches.

#### **Definitions of Assessment**

Many authors have defined “assessment” as it pertains to higher education. Popham views assessment as “educational measurement” and defines it as “an inference-making enterprise in which we formally collect overt, test-based evidence from students to arrive at what we hope are accurate inferences about students’ status with respect to covert, educationally important variables.”<sup>127</sup> Farmer states, “Assessment...refers to gathering and analyzing information about students in order to determine their abilities. Assessment is a basic part of education because it enables one to determine whether students

---

<sup>127</sup> Popham, Test Better, Teach Better. 4.



‘get it,’ whether they meet the standard.”<sup>128</sup> Angelo provides a more thorough definition of this type of assessment:

Assessment is an ongoing process aimed at understanding and improving student learning. It involves making our expectations explicit and public; setting appropriate criteria and high standards for learning quality; systematically gathering, analyzing, and interpreting evidence to determine how well performance matches those expectations and standards; and using the resulting information to document, explain, and improve performance.<sup>129</sup>

Bresciani adds to this definition six qualities that assessment should embody.

First, assessment should be meaningful and driven by the experts in the area to be assessed, such as faculty or librarians. Second, assessment should be manageable and make use of available resources. Assessment should also be flexible, objective, and ethical. Fourth, assessment should result in data that can be used for continuous improvement. Finally, assessment should promote a “culture of accountability, learning, and improvement.”<sup>130</sup>

A few authors have narrowed their definitions of assessment to the context of program evaluation. Palomba and Banta conceptualize assessment as a way to improve a program as well as student learning and development.<sup>131</sup> Bresciani describes assessment that centers on program review. She writes, “Assessment attempts to answer certain questions about a program’s intended outcomes: What are we trying to do and why are we doing it? What do we

---

<sup>128</sup> Lesley S. Johnson Farmer, "Authentic Assessment of Information Literacy through Electronic Products," Book Report 16 (1997).

<sup>129</sup> Angelo, "Reassessing (and Defining) Assessment."

<sup>130</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 18.

<sup>131</sup> Catherine A. Palomba and Trudy W. Banta, Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education (San Francisco: Jossey-Bass, 1999).

expect the students to know or do as a result of our program? How well are we doing it? How do we know? How do we use the information to improve? Does that work?"<sup>132</sup> Dugan and Hernon focus this emphasis on program assessment on the library. In describing assessment of library programs, they write:

Assessment measures changes in library users as a result of their contact with an academic library's programs, resources, and services, such as student known content, developed skills and abilities, and acquired attitudes and values. Therefore, assessment is comprised of statements about what students will know/think/be able to do as a result of their contact with library programs, not statements about what the library should/could do to bring about desired outcomes.<sup>133</sup>

In addition to general definitions of assessment, major categories of assessment exist. Pausch and Popp note that assessment is "multidimensional." They list qualitative and quantitative, formative and summative as major categories of assessment.<sup>134</sup> Quantitative assessment techniques provide data in numerical form, while qualitative data is more descriptive and usually reported in words.<sup>135</sup> Formative assessment "deals with programs as they are functioning and tries to foster ongoing improvement for the provider and receiver of the instruction."<sup>136</sup> This approach offers the advantage of making changes during,

---

<sup>132</sup> Marilee J. Bresciani, Outcomes Assessment in Student Affairs: Moving Beyond Satisfaction to Student Learning and Development, 2002, National Association of Student Personnel Administrators, 19 Feb. 2005 <<http://www.naspa.org/membership/mem/nr/article.cfm?id=884>>.

<sup>133</sup> Dugan and Hernon, "Outcomes Assessment." 378.

<sup>134</sup> Pausch and Popp, "Assessment of Information Literacy."

<sup>135</sup> Grassian and Kaplowitz, Information Literacy Instruction.

<sup>136</sup> Pausch and Popp, "Assessment of Information Literacy."

rather than after, the delivery of instruction.<sup>137</sup> Summative assessment, on the other hand, takes place after instruction is complete.<sup>138</sup>

### **Purposes of Assessment**

In 1998, the Progress Report on Information Literacy called for researchers to “benchmark information literacy skills” and “measure the effectiveness of information literacy programs on student performance”.<sup>139</sup> Librarians have responded to this and other calls for the evaluation of information literacy skills and programs by learning about and testing assessment approaches. Before librarians launch into detailed assessment efforts, however, they should carefully consider the purpose of such assessments. Shepard suggests:

The intended use of an assessment—its purpose—determines every other aspect of how the assessment is conducted. Purpose determines the content of the assessment (What should be measured?); methods of data collection (Should the procedures be standardized? Should data come from all students or from a sample of students?); technical requirements of the assessment (What level of reliability and validity must be established?); and finally, the stakes or consequences of the assessment, which in turn determine the kinds of safeguards necessary to protect against potential harm from fallible assessment-based decisions.<sup>140</sup>

Librarians seeking to assess information literacy skills and programs may have multiple purposes in mind. Among them are two overarching goals: 1) to

---

<sup>137</sup> Palomba and Banta, Assessment Essentials.

<sup>138</sup> Pausch and Popp, "Assessment of Information Literacy."

<sup>139</sup> Thompson, "Information Literacy Accreditation Mandates."

<sup>140</sup> Lorrie A. Shepard, "Standardized Tests and High-Stakes Assessment," Encyclopedia of Education, ed. James W. Guthrie, vol. 7 (New York: Macmillan Reference USA, 2003).

respond to accountability measures, including accreditation mandates, and 2) to support library instruction programs and build a culture of continuous improvement of student learning.

### ***Responding to Calls for Accountability***

Perhaps the most common purpose for assessment is to respond to a call for accountability. Frequently, assessment in higher education is prompted by requests from government officials, institutional administrators, and accreditors.<sup>141</sup> Miller reports that higher learning programs are increasingly required to provide evidence of the impact of the program on student learning and development.<sup>142</sup> He states, "Today, institutional effectiveness has become a primary focus in virtually all aspects of higher education. What was once referred to somewhat casually as 'being accountable' has evolved into an institutional way of life in most colleges and universities."<sup>143</sup>

While some administrators reject calls for accountability as potential attacks on their programs, Hoyler argues that requests for accountability measures are legitimate.<sup>144</sup> He writes:

A college degree represents the collective efforts of many faculty and staff, spanning at least four years of a student's full-time enrollment. Colleges and universities make claims about this four-year experience, and society regards the degree—not the individual courses—as a significant credential. Therefore, students, parents, trustees, politicians,

---

<sup>141</sup> Pausch and Popp, "Assessment of Information Literacy."

<sup>142</sup> Miller, Using Professional Standards for Program Assessment.

<sup>143</sup> Miller, Using Professional Standards for Program Assessment.

<sup>144</sup> Robert Hoyler, "The Road Not Taken," Change 30.5 (1998).

and society at large have a right to expect that the results will be what we claim they will be. What is more, those we serve have a right to expect that we will be interested in assessment. We have insisted on the usual rights of a profession to set its own standards and ensure the quality of its work. We have also presented our profession as one of integrity, which is committed to honesty, fairness, and objectivity in the pursuit of truth and to nurturing the young in these ideals. What we have created, then, is the legitimate expectation that we are committed to achieving our stated educational aims, anxious to improve our collective efforts to do so, and ready and willing to give account of our efforts and results. In short, outcomes assessment should be a central part of our professional ethics.<sup>145</sup>

This perspective suggests that program administrators should welcome accountability measures and the assessment projects that accompany them as opportunities to demonstrate commitment to student learning and development. Administrators that conduct assessment “cannot only speak with confidence about ‘what [they] do’... [they] can also discuss ‘how well [they] do it.’”<sup>146</sup>

### ***Participating in Accreditation Processes***

Threat of an impending accreditation review often spurs institutional and program administrators to initiate assessment projects. In fact, meeting accreditation requirements is a main purpose for assessment.<sup>147</sup> In recent years, accreditors require an increasing number of accountability measures to demonstrate student learning and to ensure that students gain the skills they need to be successful after graduation.<sup>148</sup> Institutional and program

---

<sup>145</sup> Hoyler, "The Road Not Taken."

<sup>146</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 1.

<sup>147</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 17.

<sup>148</sup> Thompson, "Information Literacy Accreditation Mandates." 220.

administrators are not the only ones to note an increase in such measures. Librarians also cite mounting attention paid by accreditation teams to libraries, library programs, and information literacy skills instruction.

Accreditors en masse acknowledge the importance of information literacy skills. According to Thompson, "Regional accreditation agencies are now stating outright that regular library instruction should be an essential part of higher education and that more educational standards call for information literacy to become a central core set of skills required for an undergraduate degree."<sup>149</sup> Gratch Lindauer states that in recent years most accreditation standards have increased their emphasis on the teaching role of libraries.<sup>150</sup> She found that many accrediting agencies connect library and information technology to student learning.<sup>151</sup> Some also acknowledge the library's role in helping students to gain information literacy skills, and a few link the library to "the quality of the learning environment."<sup>152</sup> Owusu-Ansah points out that the accreditors' call for the library's "direct participation in the education of students forced many libraries to reevaluate and address their place in the educational process."<sup>153</sup> Thompson argues that, "Libraries are no longer seen, if they ever were, as isolated agencies

---

<sup>149</sup> Thompson, "Information Literacy Accreditation Mandates." 233.

<sup>150</sup> Bonnie Gratch Lindauer, "Comparing the Regional Accreditation Standards: Outcomes Assessment and Other Trends," *Journal of Academic Librarianship* 28.1 (2002). 16.

<sup>151</sup> Gratch Lindauer, "Comparing the Regional Accreditation Standards." 16.

<sup>152</sup> Gratch Lindauer, "Comparing the Regional Accreditation Standards." 18.

<sup>153</sup> Owusu-Ansah, "Information Literacy and Higher Education." 8.

separate and apart from the major teaching and learning activities.”<sup>154</sup> These authors confirm that librarians increasingly find themselves at the front and center of accreditation processes.

The recent attention accreditation has brought to libraries offers librarians an opportunity to “stimulate the fundamental rethinking of the mission, role, and operation of the library.”<sup>155</sup> Rethinking the mission, role, and operation of the library may allow librarians to cultivate a “culture of evidence” in library organizations. Accreditation brings librarians the opportunity to reflect and “determine whether the library is asking the right questions, collecting useful data, analyzing the data effectively, disseminating the data to those who can benefit, and relying upon data effectively for decision making and improvement.”<sup>156</sup> All these activities are part of a culture of evidence. When such a culture exists, decisions are made based on facts and data, rather than anecdotal evidence, and improvements are made more rapidly and in a more directed way.

Rethinking the mission, role, and operation of the library also provides librarians an opportunity to reflect on what heretofore were “givens” about the profession. Some authors believe that reflection initiated by accreditation may result in a significant paradigm shift in librarianship. Thompson suggests that accreditation requirements demand:

---

<sup>154</sup> Thompson, "Information Literacy Accreditation Mandates." 220.

<sup>155</sup> Ralph A. Wolff, "Using the Accreditation Process to Transform the Mission of the Library," New Directions for Higher Education 90 (1995). 77.

<sup>156</sup> Wolff, "Using the Accreditation Process." 80.

A shift in the established library instruction paradigm at many institutions. Responsibility shifts from librarians teaching students how to locate materials for particular assignments, to faculty and librarians working together to embed the teaching and learning of information literacy skills systematically into syllabi and curricula. The new paradigm requires librarians and faculty to adopt a broader sense of the role of information literacy skills in higher education and in the preparation for the professional workforce. It also demands the learning of new methods and concepts by both teaching faculty and librarians, as they develop a collaborative approach to the integration of information literacy into general education and disciplinary education.<sup>157</sup>

Although accreditation processes hold the promise of rethinking and reflection among librarians, they may also cause a number of difficulties.

In many universities, the “teaching-learning role of academic libraries is well established, as are the expectations of accreditation agencies that libraries connect their evaluation of collections, resources, and services to educational outcomes.”<sup>158</sup> On other campuses, these library roles and expectations are new. As librarians reposition themselves at the center of teaching and learning (and thus accreditation) they may turn to existing measures of library effectiveness for help. Unfortunately, much of the data recorded in annual reports and reviews will not help libraries demonstrate how libraries impact student learning.<sup>159</sup> Accreditation documents do not generally outline the ways in which libraries might demonstrate the effects of their programs on institutional effectiveness.<sup>160</sup> One exception, noted by Ratterau, comes from the Middle States accreditation

---

<sup>157</sup> Thompson, "Information Literacy Accreditation Mandates." 218.

<sup>158</sup> Bonnie Gratch Lindauer, "Defining and Measuring the Library's Impact on Campuswide Outcomes," College and Research Libraries 59.6 (1998). 549.

<sup>159</sup> Gratch Lindauer, "Defining and Measuring the Library's Impact." 546-547.

<sup>160</sup> Gratch Lindauer, "Comparing the Regional Accreditation Standards." 20.



documentation. In Standard 11, Middle States suggests that “institutions may be able to provide ‘evidence of information literacy incorporated in the curriculum, with syllabi or other materials appropriate to the mode of teaching and learning, describing expectations for students’ demonstration of information literacy skills.’”<sup>161</sup> Still, universities in other regions appear confused “as accrediting bodies set standards without specific prescriptions for their achievement...librarians ponder their mandate but concede a lack of authority or resources to succeed with anything ambitious.”<sup>162</sup>

Despite the ambiguity of accreditation standards documents, two points seem clear. First, every program within a university, including library programs, should measure their progress toward goals they set for themselves.<sup>163</sup> Second, performance indicators must be developed that “demonstrate the library’s impact on desired educational outcomes and methods [must be determined] to measure them.”<sup>164</sup> Both of these goals can be more easily achieved if librarians join with their faculty counterparts to set and measure goals and performance indicators. Indeed, librarians need faculty assistance in writing goals and performance indicators, and faculty need librarian assistance in determining how to include information skills in the goals and performance indicators for academic majors.

---

<sup>161</sup> Oswald M. T. Ratterau, "Information Literacy in Self-Study and Accreditation," Journal of Academic Librarianship 28.6 (2002). 372.

<sup>162</sup> Owusu-Ansah, "Information Literacy and Higher Education." 6.

<sup>163</sup> Gratch Lindauer, "Comparing the Regional Accreditation Standards." 20.

<sup>164</sup> Gratch Lindauer, "Defining and Measuring the Library's Impact." 548.

These shared needs provide impetus for new campuswide collaborations.

Owusu-Ansah writes:

As information literacy permeates the discourse in colleges and universities in the United States, and accrediting bodies emphasize its necessity, academic libraries prepare to assume duties their own professional associations brought to the attention of a nation seeking new results in higher education. The institutions these libraries serve, whether they see an actual need or simply respond to political expediency, seek to address concerns that calls for information literacy have generated. Many turn to their libraries for suggestions and solutions. Academic librarians, hardly ever on an equal footing with classroom faculty, suddenly find themselves compelled to address issues and provide solutions that may go beyond their library walls and even infiltrate the domain of subject faculty.<sup>165</sup>

Other authors concur. Wolff notes that accreditation gives institutional administrators an opportunity to bring librarians and faculty together and ensure that library assessment is incorporated into campuswide assessment activities.<sup>166</sup>

Thompson notes, "2001 mandates from regional accreditation agencies and trends in higher education clearly call for academic librarians to serve on curriculum committees, to speak out on the direction of general education and disciplinary education, to be involved through liaison programs with the development and revision of courses, to work with faculty on exercises and assignments to improve student learning, and to assess student outcomes."<sup>167</sup>

Finally, Middle States Standard 11 suggests that there should be "collaboration

---

<sup>165</sup> Owusu-Ansah, "Information Literacy and Higher Education." 3.

<sup>166</sup> Wolff, "Using the Accreditation Process." 77.

<sup>167</sup> Thompson, "Information Literacy Accreditation Mandates." 232.

between professional library staff and faculty in teaching and fostering information literacy skills relevant to the curriculum.”<sup>168</sup>

Collaboration of librarians and faculty serves not only the best interests of the campus in accreditation processes and students and faculty in the classroom, it is also critical for the library and its role on campus. In a recent publication, Kuh and Gonyea state that, “relatively little is known about what and how students’ academic library experiences contribute to the desired outcomes of college.”<sup>169</sup> They suggest that, “library use does not appear to contribute directly to gains in information literacy...or to what students gain overall from college.”<sup>170</sup> Fortunately, Kuh and Gonyea recommend further investigation of library effectiveness and call for baseline measures of student information literacy skills.<sup>171</sup> Even so, Kuh and Gonyea’s article underscores the need for librarians to convey the importance of the library and information literacy skills to the university as a whole.

Gratch Lindauer confirms that librarians are key players on campus and must document and measure the ways in which the library impacts the “academic quality of life” on campus.<sup>172</sup> In fact, she argues that “the assessment of library performance should be defined and shaped by its connections and

---

<sup>168</sup> Ratterau, "Information Literacy in Self-Study and Accreditation." 372.

<sup>169</sup> George D. Kuh and Robert M. Gonyea, "The Role of the Academic Library in Promoting Student Engagement in Learning," College and Research Libraries 64.4 (2003). 258.

<sup>170</sup> Kuh and Gonyea, "The Role of the Academic Library." 267, 270.

<sup>171</sup> Kuh and Gonyea, "The Role of the Academic Library." 267-268.

<sup>172</sup> Gratch Lindauer, "Defining and Measuring the Library's Impact." 546.

contributions to institutional goals and desired educational outcomes.”<sup>173</sup> Gratch Lindauer reminds readers that information literacy skills are directly tied to general education outcomes including critical thinking, technical literacy, problem-solving, and lifelong learning.<sup>174</sup> This approach is supported by many accreditation agencies, professional organizations in the academic disciplines that focus on student competencies, and funding organizations.<sup>175</sup> It is also supported in the library literature, which recognizes that assessment, particularly assessment in response to national mandates, lends credibility to the library and attention campuswide.<sup>176</sup>

### ***Improving Program Performance***

Some significant purposes for assessment are to 1) inform decision-making<sup>177</sup> in order to 2) improve the quality or performance of a program<sup>178</sup> and 3) build from a series of improvements to a “culture of continuous improvement.”<sup>179</sup> These purposes are significant, even in the absence of calls for accountability. Miller writes, “Self-regulation constitutes a commitment to

---

<sup>173</sup> Gratch Lindauer, "Defining and Measuring the Library's Impact." 547.

<sup>174</sup> Gratch Lindauer, "Defining and Measuring the Library's Impact." 549.

<sup>175</sup> Thompson, "Information Literacy Accreditation Mandates." 237-238.

<sup>176</sup> Grassian and Kaplowitz, Information Literacy Instruction. 266.

<sup>177</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 16.

<sup>178</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 16

<sup>179</sup> Maki qtd. in Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 18.

improving program practice and organizational health.”<sup>180</sup> Commitment to improvement is particularly relevant for library instruction programs that, although not currently included in accreditation examinations, require reflection and change to increase student learning.

Several authors confirm the role of assessment in program improvement. Samson writes, “Assessment offers a value-added dimension to a library instruction program. It provides a beginning point to ascertain the program’s effectiveness and to guide direction for future instruction.”<sup>181</sup> Samson also acknowledges Iannuzzi’s claim that assessment is “essential” to the development of library instruction programs.<sup>182</sup> Without assessment, program weaknesses cannot be identified and corrected through effective decision-making. Knight acknowledges the importance of documenting both the strengths and “areas of improvement” in library instruction programs.<sup>183</sup> Miller describes the importance of finding and confronting program weaknesses as “essential”.<sup>184</sup> Perhaps Barclay states the connection between assessment and improvement most clearly: “Unless evaluation will somehow improve the thing being evaluated, it is not worth doing.”<sup>185</sup>

---

<sup>180</sup> Miller, Using Professional Standards for Program Assessment..

<sup>181</sup> Samson, "What and When Do They Know? Web-Based Assessment."

<sup>182</sup> Samson, "What and When Do They Know? Web-Based Assessment."

<sup>183</sup> Knight, "The Role of Assessment in Library User Education."

<sup>184</sup> Miller, Using Professional Standards for Program Assessment..

<sup>185</sup> Donald Barclay, "Evaluating Library Instruction: Doing the Best You Can with What You Have," Research Quarterly 33.2 (1993).

## ***Improving Program Structures***

In addition to the greater goal of improving program performance, assessment can have the very practical purpose of improving the structure of a program. For example, assessment can be used to connect a program mission with the missions of the larger organizational body or institution. Wolff believes that “the mission of the library of the future will be principally a teaching one that is directly linked to the educational mission of the institution. Thus, assessment should be directed primarily at the library’s relationship to the teaching and learning functions of the institution.”<sup>186</sup> He also states that the “evaluation of the library needs to include...the library’s relationship to the mission of the institution [and] the effectiveness of the library in accomplishing its mission.”<sup>187</sup>

Assessment can also be used to reinforce, emphasize, align or realign activities with a previously defined mission of a program.<sup>188</sup>

In addition to checking and aligning program missions, assessment can be used to celebrate successes.<sup>189</sup> Success can be leveraged to form collaborations, find funding, and champion change. In fact, assessment both requires and generates collaboration between librarians and university faculty. ACRL encourages “future research in the areas of assessment, evaluation, and transferability [of information literacy skills]...to address involvement from

---

<sup>186</sup> Wolff, "Using the Accreditation Process." 90.

<sup>187</sup> Wolff, "Using the Accreditation Process." 78.

<sup>188</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 16.

<sup>189</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 17.

stakeholders other than librarians.”<sup>190</sup> Carter suggests that assessment projects in the library help librarians develop more projects with faculty.<sup>191</sup> Knight asserts that the data resulting from assessments help strengthen librarian connections with faculty members.<sup>192</sup>

For library instruction programs, assessment can support requests to continue or increase funding. Bresciani, Zelna, and Anderson list requesting additional funds as a main purpose of assessment.<sup>193</sup> This purpose is important for information literacy instruction programs, especially those that must justify their existence or risk losing financial support.<sup>194</sup> Pausch and Popp state that library instruction programs must demonstrate their impact on other learning experiences to gain support from library administrators, institutional administrators, faculty, and students.<sup>195</sup> Even if assessment yields negative results, it can still be used to improve program structures. According to Grassian and Kaplowitz, Corcoran and Langlois assert that negative results help “pinpoint problem areas where improvements could lead to a stronger, better program in the future.”<sup>196</sup> Grassian and Kaplowitz suggest that librarians can use negative results to argue for greater financial support by highlighting where funding could

---

<sup>190</sup> Research and Scholarship Committee of the ACRL Instruction Section, "Research Agenda for Library Instruction and Information Literacy," College and Research Libraries 64.2 (2003). 112.

<sup>191</sup> Carter, "'Doing the Best You Can with What You Have'." 41.

<sup>192</sup> Knight, "The Role of Assessment in Library User Education."

<sup>193</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 17.

<sup>194</sup> Grassian and Kaplowitz, Information Literacy Instruction. 266.

<sup>195</sup> Pausch and Popp, "Assessment of Information Literacy."

<sup>196</sup> Grassian and Kaplowitz, Information Literacy Instruction. 266.

be used to improve the program.<sup>197</sup> Assessment projects that lead to increased funding serves the ultimate purpose of assessment—change. As Knight suggests, “It is...important to view assessment programs not as ends...but ...as significant sources of information that foster feedback for change.”<sup>198</sup>

### ***Improving Student Learning and Teacher Skills***

Although assessment is often conducted to respond to calls for accountability, participate in accreditation, or improve programs, many authors argue that the main purpose of assessment is to improve student learning. According to Popham, “The central mission of all...assessment is (1) to help you make valid inferences about your students so that you can then (2) make better decisions about how to instruct those students.”<sup>199</sup> Institutions need to conduct assessment projects because they supply the data required to make defensible decisions and ultimately achieve quality programs, and the direct evidence assessment provides is clearly linked to indicators of institutional quality.<sup>200</sup>

According to Hanson’s foreword to Assessing Student Learning and Development: A Handbook for Practitioners:

Quality is on everyone's mind in higher education. Students want it, parents pay for it, and corporate businesses and industries demand it.... An important facet of this demand for quality is the expectation that students learn important knowledge, skills, attitudes, and behaviors as a result of attending college. Colleges and universities can no longer count

---

<sup>197</sup> Grassian and Kaplowitz, Information Literacy Instruction. 266.

<sup>198</sup> Knight, "The Role of Assessment in Library User Education."

<sup>199</sup> Popham, Test Better, Teach Better. 69.

<sup>200</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 1.



the number of books in the library, brag about the average SAT or ACT score of the most recent entering class, or simply report the student-faculty ratio and the six-year graduation rate. The focus has instead become, "What have students learned?" and "What are they able to do with what they have learned?"<sup>201</sup>

Programs that comprise institutions of higher education, including library instruction programs, must demonstrate the quality that students, parents, and employers seek.

Assessment allows programs to demonstrate their contribution to student learning and a quality university experience. The ACRL Instruction Section recognizes that library instruction programs must become involved in assessment to document the campuswide impact of library instruction and information literacy.<sup>202</sup> Instruction librarians must "find out whether or not what is taught is useful."<sup>203</sup> They must demonstrate that students acquire information literacy skills through library instruction.

Assessment not only documents student learning, but also provides important feedback teachers can use to improve their skills. Grassian and Kaplowitz state that librarians, "assess, evaluate, and revise because we want to find out if our instruction has been effective. In other words, we need to find out how well our goals and objectives have been met. Furthermore, we want to highlight areas where our efforts might be improved for the future. Developing

---

<sup>201</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. vii.

<sup>202</sup> Research and Scholarship Committee of the ACRL Instruction Section, "Research Agenda." 112.

<sup>203</sup> Pausch and Popp, "Assessment of Information Literacy." "

instruction is an iterative process."<sup>204</sup> Librarians can use assessment results to alter and improve instruction.<sup>205</sup> They can use results to learn about what students already know, what to teach, how long to teach it, and how effective the instruction is.<sup>206</sup> Librarians can also use assessment to reflect on their teaching,<sup>207</sup> their attitudes and approaches to learning,<sup>208</sup> and their assumptions about learning.<sup>209</sup> Sampson writes, "Assessment provides the opportunity to take a fresh look at the classroom experience."<sup>210</sup> Carter notes that assessment teaches librarians "not to assume anything when dealing with students."<sup>211</sup> In addition, assessment provides opportunities for teaching librarians to extend their content knowledge about learning and assessment. In fact, assessment teaches librarians about the learning process itself.<sup>212</sup> Arter confirms that assessment "demystif[ies] the learning process" and makes teachers and learners equal stakeholders in learning.<sup>213</sup> Furthermore, teachers and students can be partners

---

<sup>204</sup> Grassian and Kaplowitz, Information Literacy Instruction. 265.

<sup>205</sup> Knight, "The Role of Assessment in Library User Education."

<sup>206</sup> Popham, Test Better, Teach Better. 5-6.

<sup>207</sup> Dorothy Anne Warner, "Programmatic Assessment: Turning Process into Practice by Teaching for Learning," Journal of Academic Librarianship 29.3 (2003). 169.

<sup>208</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 17.

<sup>209</sup> Warner, "Programmatic Assessment." 169.

<sup>210</sup> Samson, "What and When Do They Know? Web-Based Assessment."

<sup>211</sup> Carter, "'Doing the Best You Can with What You Have'." 40.

<sup>212</sup> Grassian and Kaplowitz, Information Literacy Instruction. 266.

<sup>213</sup> Judith A. Arter, "Using Assessment as a Tool for Learning," A Handbook for Student Performance in an Era of Restructuring, eds. Robert E. Blum and Judith A. Arter (Alexandria, Virginia: Association for Supervision and Curriculum Development, 1996). IV-10:1.

in the assessment process. Shepard describes teachers and students as equal practitioners of assessment in a community of learners:

I believe it will be helpful for teachers to make their investigations of teaching visible to students, for example, by discussing with them decisions to redirect instruction, stop for a mini-lesson, and so forth.... If we want to develop a community of learners—where students naturally seek feedback and critique their own work—then it is reasonable that teachers would model this same commitment to using data systematically as it applies to their own role in the teaching and learning process.<sup>214</sup>

Working together, students and librarians can improve not just instruction, but also assessment methods. Hanson writes, “That assessment should be a high priority is no longer a point of discussion.... The more challenging question is how to do assessment well.”<sup>215</sup> By striving to do assessment well, librarians can use what they learn from themselves and from students to “refine the design of future assessment methods.”<sup>216</sup>

### ***Closing the Loop***

One final purpose for assessment works in tandem with the others. That purpose is to “close the loop”, a phrase attributed to Maki.<sup>217</sup> To close the loop, educators finish the assessment cycle by using data to improve teaching and learning programs and increase student learning and development.

---

<sup>214</sup> Lorrie A. Shepard, “The Role of Assessment in a Learning Culture,” Educational Researcher 29.7 (2000). 12.

<sup>215</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. vii.

<sup>216</sup> Knight, “The Role of Assessment in Library User Education.”

<sup>217</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 10.

Unfortunately, this point is lost on some librarians. For example, Hiller and Self write:

An assessment procedure may identify problems, but it will probably not tell us how to fix them. Typically, a more in-depth study is required to solve a problem. Because assessment focuses on the present, and the past, it generally does not tell us what we should be doing in the future. Typically the creative, innovative ideas, the revolutionary changes, do not come from assessment.<sup>218</sup>

Other librarians fully understand the core purpose of assessment. Carter states, "To be meaningful...assessment must collect hard data, and librarians must use that data to evaluate their programs and make changes necessary to improve those programs."<sup>219</sup> Samson states, "An assessment is only valuable when the analyses are used to augment or change the program being assessed."<sup>220</sup>

Grassian and Kaplowitz also grasp the cyclical nature of assessment and the continuing challenge to close the loop. They summarize:

We plan. We develop. We deliver. We assess and evaluate the results of the assessment. We revise, deliver the revised material, and assess and evaluate again. Perfection is always just out of reach; but continually striving for perfection contributes to keeping both our instruction fresh and our interest in teaching piqued.<sup>221</sup>

### **Selecting a Method of Assessment**

Once a decision is made to conduct an assessment, the next step is to select a method or tool for assessment. Maki writes:

---

<sup>218</sup> Steve Hiller and Jim Self, "Making Library Assessment Work," ARL 4th Human Resources Management Symposium (Washington, D.C.: 2004).

<sup>219</sup> Carter, "'Doing the Best You Can with What You Have'." 41.

<sup>220</sup> Samson, "What and When Do They Know? Web-Based Assessment."

<sup>221</sup> Grassian and Kaplowitz, Information Literacy Instruction. 265.

As institutions increasingly commit to assessing student learning...the first challenge they face is to decide which assessment methods to use. Identifying or developing assessment methods requires a thorough understanding of what each actually measures and how each relates to a program's articulated outcomes and expected levels of student performance. That is, what methods best enable an institution to determine how well students have met its expectations?<sup>222</sup>

Rather than selecting the method that is most familiar, librarians should select assessment methods intentionally.<sup>223</sup> Among the criteria librarians should consider are: utility, relevance to learning, integration into curriculum, needs of stakeholders, needs of those being assessed, psychometrics, credibility, cultural fairness, scope, and cost. In addition, best practices should be sought out and followed whenever possible.

### ***Utility and Relevance***

Before selecting an assessment method to evaluate a library instruction program, teaching librarians should determine the utility of the method. That is, will the data generated by the measure be useful in guiding subsequent decisions?<sup>224</sup> To determine this, librarians should consider what the test results will be used for,<sup>225</sup> and what instructional decisions need to be made.<sup>226</sup>

---

<sup>222</sup> Peggy L. Maki, "From Standardized Tests to Alternative Methods: Assessing Learning in Education," Change 33.2 (2001).

<sup>223</sup> Bresciani, Zelná and Anderson, Assessing Student Learning and Development. 25.

<sup>224</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment, Volume 1: Definitions and Assessment Methods for Critical Thinking, Problem Solving, and Writing (Harrisonburg, Virginia: National Center for Education Statistics, 2000). 2-3.

<sup>225</sup> Popham, Test Better, Teach Better. 61.

<sup>226</sup> Popham, Test Better, Teach Better. 61.

Bresciani suggests asking three questions to determine the utility of a measure: “Will this assessment method help me understand what it is that is contributing to the end result stated in [the measured] outcome? Will this assessment method help me understand why I am delivering the services in the way that I am? Will the evidence collected from this method help me understand how to maintain status quo?”<sup>227</sup>

In addition to utility, librarians should consider the relevance a particular assessment method has to learning and ask themselves whether or not the assessment method will answer questions about student learning. Grassian and Kaplowitz confirm that, “assessment methods should be selected because of their relevance to the learning outcomes or performances to be measured.”<sup>228</sup> Evans addresses the importance of aligning assessment measures with learning goals:

This alignment is essential because the more closely students’ educational experiences match the content covered by the assessment measure, the greater the utility of the assessment results. Educators who want to know about the effectiveness of their educational programs need to use measures that accurately represent expected outcomes.<sup>229</sup>

Prus and Johnson state unequivocally that, “If an assessment method doesn’t measure what your program teaches, or doesn’t measure it with precision, or doesn’t suggest what the program’s strengths and weaknesses are, then that assessment method cannot serve the institutional effectiveness goals of your

---

<sup>227</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 20.

<sup>228</sup> Grassian and Kaplowitz, Information Literacy Instruction. 9.

<sup>229</sup> Candy Duncan Evans, "Understanding Assessment," Delta Pi Epsilon Journal 44.1 (2002). 9.

program.”<sup>230</sup> Prus and Johnson take relevance to student learning one step further and assert the importance of integrating assessment into learning curriculum. They include “maximum incorporation into activities in the ongoing academic program” as an ideal characteristic of an assessment method.<sup>231</sup>

### ***Stakeholder Needs***

When selecting an assessment tool, librarians should consider the needs of stakeholders, including both the audience that assessment data will be reported to and the participants who will undergo assessment. Grassian and Kaplowitz recognize the importance of audience in selecting an assessment method.<sup>232</sup> Others list stakeholder needs as a key consideration in selecting an assessment measure, stating:

Applicability of assessment measures relates to the extent to which information on a particular outcome measure meets the needs of multiple stakeholder groups. In other words, to what extent will [the] data generated...yield information that can be used by multiple groups, such as faculty and administrators who wish to improve programs, or government officials and prospective employers who desire documentation of skill level achievement or attainment?<sup>233</sup>

Grassian and Kaplowitz list a number of considerations in assessment tool selection, including who will see final results, the plans for using the results, level of preciseness, level of detail, speed of reporting, and qualitative and quantitative

---

<sup>230</sup> Joseph Prus and Reid Johnson, "A Critical Review of Student Assessment Options," New Directions for Community Colleges 88 (1994). 70.

<sup>231</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 70.

<sup>232</sup> Grassian and Kaplowitz, Information Literacy Instruction. 271.

<sup>233</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment. 2-3.

preferences.<sup>234</sup> Librarians should also recognize that stakeholders need assessment data to be displayed in a comprehensible format.<sup>235</sup> Assessment measures also need to evaluate both program outcomes and individual student progress.<sup>236</sup> Finally, librarians undertaking assessment should reflect upon the needs of those who are assessed, including how many there are, who they are, and what their assessment experience might be.<sup>237</sup>

### ***Measurability and Cost***

Two additional considerations in choosing a method for assessment are measurability and cost. "Measurability refers to how [an] outcome is operationally defined and measured, including the methodological soundness of the chosen measure."<sup>238</sup> Colton, et al. defines measurability as whether or not the tool measures what it intends to measure (validity) with consistency (reliability), and he lists measurability as the first focus in choosing an assessment tool.<sup>239</sup> Other sources also list psychometric properties such as

---

<sup>234</sup> Grassian and Kaplowitz, Information Literacy Instruction. 272.

<sup>235</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment. 2-3.

<sup>236</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 70.

<sup>237</sup> Assessment Methods, 1999, Assessment Strategies Inc.  
<<http://www.asinc.ca/en/methods/index.html>>.

<sup>238</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment. 3-4.

<sup>239</sup> Dean A. Colton, Xiaohong Gao, Deborah J. Harris, Michael J. Kolen, Dara Martinovich-Barhite, Tianyou Wang and Catherine J. Welch, Reliability Issues with Performance Assessments: A Collection of Papers (Iowa City: ACT, 1997). 6



validity and reliability as key factors in selecting an assessment method.<sup>240</sup> Prus and Johnson add internal validity defined as “maximum relevance to the unique aspects of the local program curriculum” and external validity defined as “maximum generalizability to similar programs at colleges across the state, region, and nation.”<sup>241</sup> All of these aspects of measurability can impact the credibility of an assessment method.<sup>242</sup>

In addition to psychometric properties, a number of other criteria establish the credibility level of an assessment measure. These include: “the amount of time, energy, and expertise that goes into a particular measure...the ease of interpretation of the materials and results; the amount of detail provided pertaining to student outcomes...the conceptual [relationship] to the actual skills deemed important...and the cultural fairness.”<sup>243</sup> The last of these criteria, cultural fairness, means that the assessment method should “not be biased or misleading in favor of particular groups.”<sup>244</sup>

A final key consideration in selecting an assessment method is cost.<sup>245</sup> Ideally, educators should choose the assessment method that best fits their

---

<sup>240</sup> Assessment Methods.

<sup>241</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 70.

<sup>242</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment. 2-3.

<sup>243</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment. 2-3.

<sup>244</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment. 2-3.

<sup>245</sup> Assessment Methods.

programmatic and instructional needs, but cost is a practical concern that cannot be ignored. Bresciani, Zelna and Anderson acknowledge:

Budget constraints are a reality. Higher education is facing major budget cuts in many states. It is often difficult to find funding for new projects, and assessment activities may fall into a 'new' project category. When determining methods and instruments, you should consider the costs. Not only should you consider the cost of the tool, but also the cost of implementation and analysis.<sup>246</sup>

Cost is impacted first by the scope of an assessment and whether entire populations must be assessed or a sample will suffice. According to the National Postsecondary Education Cooperative Student Outcomes Pilot Working Group:

If "census-type" data drawn from all students in attendance...are needed, then researchers should opt for measures that can be efficiently administered and scored.... However, if the scope of data needed is more restricted...with examinees selected via sampling strategies requiring fewer participants...then measures designed to assess more highly-specified curriculum based skills can be used.... For the purposes of accountability, it is not necessary to assess every student to derive valid estimates of system performance, and a much wider range of outcome data can be generated when careful sampling is conducted.<sup>247</sup>

Time is another expense that should be considered. In fact, time is one of the most common reasons for not conducting assessment at all.<sup>248</sup> Librarians should consider the time that is required for creating a tool, implementing it and analyzing it.<sup>249</sup> Prus and Johnson agree that librarians should consider whether

---

<sup>246</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 26.

<sup>247</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment.3-4.

<sup>248</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 26.

<sup>249</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 26.

the cost of a particular assessment method is practical, including time, effort, and money in their deliberations.<sup>250</sup>

### ***Best Practices***

When selecting a method of assessment, librarians should strive to follow best practices whenever possible. Among the best practices to consider are matching qualitative vs. quantitative tools to assessment needs, using multiple methods, practicing continuous assessment, beginning with a pilot test, and striving for practical methodologies that can be integrated into normal program workflow.

Both qualitative and quantitative assessment methods can produce useful results. The key is to match assessment methods with program needs, especially the needs of stakeholders who will view the information that results from the assessment. Hiller and Self submit that, "in a serious assessment, it is often good to have both quantitative results...and qualitative information."<sup>251</sup> Others note that while some stakeholders may prefer numeric results, "qualitative assessment is to be preferred when the nature of the program or its aims can best (or adequately) measured without the use of numbers or when detailed descriptions provide the best data for altering, judging or continuing, a program."<sup>252</sup> Grassian and Kaplowitz remind that "the choice of using

---

<sup>250</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 70.

<sup>251</sup> Hiller and Self, "Making Library Assessment Work."

<sup>252</sup> Grassian and Kaplowitz, Information Literacy Instruction. 277.

quantitative or qualitative data does not reflect on the rigor of the assessment.”<sup>253</sup>

Thus, matching qualitative or quantitative methods to the needs of stakeholders is the best policy to follow when selecting an assessment method.

A second best practice to consider is the use of multiple methods in assessment. Maki states that good practice includes methods ranging from standardized tests to alternative methods.<sup>254</sup> Using a variety of methods acknowledges that “different assessment methods have particular strengths for assessing some knowledge, skills and abilities, but are not ideal for others.”<sup>255</sup> By using multiple methods, librarians gain a variety of assessment feedback. Also, this approach allows librarians “to obtain maximum validity and to reduce potential error or bias associated with any one approach.”<sup>256</sup>

Practicing continuous, rather than “one-shot,” assessments is necessary to produce useful assessment data. Grassian and Kaplowitz recognize that assessments conducted solely for the purpose of accountability are usually performed at regular intervals. However, they emphasize that assessments for reflection and improvement should be ongoing.<sup>257</sup> In fact, the need for ongoing assessment is outlined in the AAHE principles of good practice:

Assessment works best when it is ongoing not episodic. Assessment is a process whose power is cumulative. Though isolated, “one-shot” assessment can be better than none, improvement is best fostered when

---

<sup>253</sup> Grassian and Kaplowitz, Information Literacy Instruction. 277.

<sup>254</sup> Maki, "From Standardized Tests to Alternative Methods."

<sup>255</sup> Assessment Methods.

<sup>256</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 83.

<sup>257</sup> Grassian and Kaplowitz, Information Literacy Instruction. 271.

assessment entails a linked series of activities undertaken over time. This may mean tracking the process of individual students, or of cohorts of students; it may mean collecting the same examples of student performance or using the same instrument semester after semester. The point is to monitor progress toward intended goals in a spirit of continuous improvement. Along the way, the assessment process itself should be evaluated and refined in light of emerging insights.<sup>258</sup>

Pilot testing is also considered a best practice when selecting an assessment method. Prus and Johnson contend, “the only way to be certain that a particular methodological option is good for your program is to pilot-test it on your students, in your curriculum, with your faculty—an educated trial-and-error approach.”<sup>259</sup> Grassian and Kaplowitz tell librarians that after selecting an assessment approach and adapting it to their particular program, field-testing is required. They write, “Make sure that you try your methods out on a sample group that as closely approximates your target population as possible.”<sup>260</sup> Another approach is to treat the first deployment of an assessment method as a pilot test and then to compare the results to other assessment methods that measure the same thing.<sup>261</sup> This approach combines the needs for pilot testing and the use of multiple methods.

Librarians should strive to include these best practices into the assessment of library instruction, but practical constraints should also be acknowledged. Grassian and Kaplowitz recognize potential difficulties in planning perfect information literacy assessments:

---

<sup>258</sup> American Association of Higher Education, Nine Principles.

<sup>259</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 70.

<sup>260</sup> Grassian and Kaplowitz, Information Literacy Instruction. 286.

<sup>261</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 83.

In choosing an assessment technique, there is the educational ideal...and then there are the practical constraints.... [information literacy] instructors operate under special circumstances. With the exception of full semester [information literacy] courses, we generally have little or no authority over our learners, have limited contact time with them, may never get the opportunity to see how they might perform in real life, and probably won't have the chance to examine the products of their learning. So most [information literacy instruction] assessments must occur on the spot in parallel with the instruction itself.<sup>262</sup>

Other authors acknowledge a need for practicality as well. The National Postsecondary Education Cooperative Student Outcomes Pilot Working Group lists "availability" as a key factor to consider in choosing an assessment method. In this context, availability "refers to issues revolving around the availability of existing measures, the feasibility of developing new measures, and the logistics of using specified measures."<sup>263</sup> The availability of existing measures for assessing library instruction is a particularly difficult issue. In the information literacy assessment field, "trying to find an already developed instrument that will suit your specific needs may not be easy. While it is comforting to use an instrument that has already gone through some field testing for reliability, validity, and usability, it may be difficult to find one that really suits your purpose."<sup>264</sup> Prus and Johnson recommend balance and reflection in choosing a practical assessment method. They suggest:

Search for ideal methods, but recognize that the ideal usually means methods that are the best fit between program needs, satisfactory validity, and affordability. Recognize that development of an assessment program

---

<sup>262</sup> Grassian and Kaplowitz, Information Literacy Instruction. 272.

<sup>263</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment. 3-4.

<sup>264</sup> Grassian and Kaplowitz, Information Literacy Instruction. 280.

is a dynamic process. Ongoing assessment of assessment methods themselves is an important part of that process.<sup>265</sup>

Although selection of an assessment method may seem a daunting task, it is one that can be guided by criteria for selection and best practices. Librarians attempting to select assessment methods must understand the strengths and weaknesses of various techniques.<sup>266</sup> They also must consider the types of information that various techniques provide, and whether or not the methods they select will answer the questions they have about their library instruction programs.<sup>267</sup> Ultimately, the choice of assessment comes down to a fit between the purposes of assessment and the capabilities of assessment methods. As Grassian and Kaplowitz point out, "An assessment method is neither good nor bad in its own right. The problem arises when an inappropriate technique is used to answer a particular assessment question."<sup>268</sup>

### **Assessment of Library Instruction: The Past**

Library instruction literature includes a limited number of works about assessment. The absence of library instruction assessment research has been remarked upon for the last quarter century. In 1980, Werking noted that library instruction programs were not conducting meaningful evaluations.<sup>269</sup> In 1981, an

---

<sup>265</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 83.

<sup>266</sup> Grassian and Kaplowitz, Information Literacy Instruction. 279.

<sup>267</sup> Grassian and Kaplowitz, Information Literacy Instruction. 273.

<sup>268</sup> Grassian and Kaplowitz, Information Literacy Instruction.

<sup>269</sup> Richard Hume Werking, "Evaluating Bibliographic Education: A Review and Critique," Library Trends 29.1 (1980).

editorial in *College and Research Libraries* acknowledged the need for assessment and the amount of work to be done. Schmidt stated, "librarians...have long been aware of the need to evaluate instruction...what is clear...is that we...have a long way to go."<sup>270</sup> Barclay noted that Hardesty, Lovich, and Mannon stated in 1982 that librarians were talking about, but not doing, evaluation of library instruction.<sup>271</sup> Barclay also cites a Cottam article from 1982 that states that any library instruction evaluation actually conducted was an "afterthought" and was not integrated into the teaching and learning process.<sup>272</sup> Pausch and Popp remind readers of a library literature study, conducted from 1980 to 1993 by Bober, Poulin, and Vilen, that asserts assessment of library instruction is "informal in nature."<sup>273</sup> Eadie has complained, according to Barclay, that too much library instruction evaluation focused on user satisfaction rather than learning,<sup>274</sup> and Mensching found in 1989 that while a meager sixty-two percent of library instruction programs conducted some form of evaluation, only twenty-three percent used testing.<sup>275</sup> In 1997, Pausch and Popp reviewed the then recent literature and found that it "reveals few changes in the formal evaluation methodologies employed by librarians."<sup>276</sup> In 2000, Grassian and

---

<sup>270</sup> C. James Schmidt, "Editorial," *College and Research Libraries* 42.1 (1981). 5.

<sup>271</sup> Barclay, "Evaluating Library Instruction."

<sup>272</sup> Barclay, "Evaluating Library Instruction."

<sup>273</sup> Pausch and Popp, "Assessment of Information Literacy."

<sup>274</sup> Barclay, "Evaluating Library Instruction."

<sup>275</sup> Teresa B. Mensching, "Trends in Bibliographic Instruction in the 1980's: A Comparison of Data from Two Surveys," *Research Strategies* 7.4 (1989).10-11.

<sup>276</sup> Pausch and Popp, "Assessment of Information Literacy."



Kaplowitz reviewed literature found in ERIC, Library Literature, and LISA and found “little evidence of any in-depth rigorous assessment studies.”<sup>277</sup> Grassian and Kaplowitz conclude, “If information professionals are engaged in these types of studies, they are not reporting it.”<sup>278</sup>

### ***Inputs and Outputs***

Historically, library instruction has been assessed using input and output measures, satisfaction surveys, or anecdotal reporting. None of these assessment approaches demonstrated student learning, and in recent years they have fallen out of favor. Iannuzzi asks, “Why is the assessment of student learning in academic libraries so elusive?”<sup>279</sup> According to Iannuzzi, one possible answer is the historically strong preference in libraries for quantitative analysis rather than qualitative analysis.<sup>280</sup> Alternatively, Ratterau blames library standards that focus on inputs and outputs.<sup>281</sup> He writes, “Librarians, faculty, and administrators became quite comfortable with this approach, routinely producing reams of data on inputs and outputs. In their analyses, data on outputs effectively became proxies for student learning.”<sup>282</sup> Dugan and Hernon also describe the measures libraries collected as being “comprised mostly of inputs

---

<sup>277</sup> Grassian and Kaplowitz, Information Literacy Instruction.

<sup>278</sup> Grassian and Kaplowitz, Information Literacy Instruction. 269-270.

<sup>279</sup> Patricia Iannuzzi, “We Are Teaching, but Are They Learning: Accountability, Productivity, and Assessment,” Journal of Academic Librarianship 25.4 (1999). 304.

<sup>280</sup> Iannuzzi, “We Are Teaching, but Are They Learning.” 304.

<sup>281</sup> Ratterau, “Information Literacy in Self-Study and Accreditation.” 369.

<sup>282</sup> Ratterau, “Information Literacy in Self-Study and Accreditation.” 369.

(resources provided to conduct activities) and outputs (the results of the inputs applied, usually quantifiably measured).<sup>283</sup> They note that this approach fails to portray libraries as an integral part of the learning process.<sup>284</sup> McDonald and Neill suggests that “circulation statistics, book counts, and other traditional measures may not be relevant because they are limited in detailing the direct impact of learning resources programs in effecting successful learning outcomes.”<sup>285</sup> Smith addresses the problem of using outputs for the assessment of library instruction when he writes, “The focus to date is primarily on making information more and more accessible rather than addressing specifically the learning outcomes important to student success.”<sup>286</sup> Knight takes the same position and laments that, “In the past, librarians relied on output statistics, such as library instruction sessions taught and number of students attending, to confirm their relevance to the institution's educational mission.”<sup>287</sup> Dugan and Hernon restate the problem boldly, “The current applications of input and output measures miss the point—if the mission of the university is teaching/learning and research, how do these descriptive library inputs and outputs measure these?”<sup>288</sup> All of these authors make the same point: the input and output measures that

---

<sup>283</sup> Dugan and Hernon, "Outcomes Assessment." 376.

<sup>284</sup> Dugan and Hernon, "Outcomes Assessment." 376.

<sup>285</sup> Marilyn McDonald and Gretchen H. Neill, "Standards for Community, Junior, and Technical College Learning Resources Programs," College and Research Libraries News 55 (1994). 572-573.

<sup>286</sup> Kenneth R. Smith, New Roles and Responsibilities for the University Library: Advancing Student Learning through Outcomes Assessment (Association of Research Libraries, 2000).

<sup>287</sup> Knight, "The Role of Assessment in Library User Education."

<sup>288</sup> Dugan and Hernon, "Outcomes Assessment." 379.

librarians have long used to assess their services do not help libraries demonstrate their contribution to student learning.

Once librarians acknowledge that input and output measures do not serve their purposes in demonstrating student learning, they must recognize that continuing to use them may have negative consequences. Ratterau suggests that libraries should no longer expect to produce such lists of inputs and outputs and should not expect their larger institutions to accept them.<sup>289</sup> Wolff cedes that input and output measures might have been effective responses to resource or research indicators in the past, but maintains that they are an ineffective response to institutional calls for demonstrations of student learning.<sup>290</sup> He continues, "Separate standards for libraries that focus on holding size, budgets, staff, and facilities no longer make sense. They must be linked to the educational purpose of the institution as a whole."<sup>291</sup> Carter agrees, especially when programs that cannot demonstrate an impact on student learning are in danger. She writes, "Annual statistics, such as how many books bought and circulated, just do not measure up in a dollar-crunching environment."<sup>292</sup>

Another major reason to move beyond input and output measures for assessing library instruction is the recent change in focus of institutional accreditation. Smith writes that, "institutional and professional accreditation bodies have been shifting their attention from input measures (faculty, courses,

---

<sup>289</sup> Ratterau, "Information Literacy in Self-Study and Accreditation." 375.

<sup>290</sup> Wolff, "Using the Accreditation Process." 79.

<sup>291</sup> Wolff, "Using the Accreditation Process." 82.

<sup>292</sup> Carter, "'Doing the Best You Can with What You Have'." 41.

books) to outcomes measures (what students learn). Universities and colleges are required to develop and implement a student outcomes assessment program."<sup>293</sup> As universities turn to assessing student learning outcomes, so must libraries. Dugan and Hernon note that, "regional accrediting bodies are increasingly focusing on determining learning results rather than counting library books."<sup>294</sup> They state that in the past accreditors accepted inputs and outputs as accountability measures, but provide examples of two regional accrediting bodies that ask for "information concerning student acquisition of information literacy skills, not just descriptive data concerning the availability and access of information resources."<sup>295</sup> Ratterau sums up accreditors' new expectations:

Instead of focusing up to two years of the self-study process on gathering and analyzing data on inputs and outputs, concluding with a voluminous self-study report, librarians and faculty now are being asked to engage in what will be for many either a new or a deeper level of collaboration. They are being asked essentially to discuss the impact of information literacy instruction, and by extension the library and librarians, on the core learning that students demonstrate as they prepare work products and other performances to fulfill the syllabi requirements of faculty. They are expected to work together to develop, implement, and assess a strategy for information literacy instruction that will lead to individual and institutional improvement.<sup>296</sup>

Clearly, input and output measures of the past will no longer answer the questions librarians, faculty, administrators, and accreditors pose about the contribution of the library and the library instruction program to student learning.

---

<sup>293</sup> Smith, New Roles and Responsibilities for the University Library.

<sup>294</sup> Dugan and Hernon, "Outcomes Assessment."

<sup>295</sup> Dugan and Hernon, "Outcomes Assessment."

<sup>296</sup> Ratterau, "Information Literacy in Self-Study and Accreditation." 370.

## ***Satisfaction Surveys***

Past assessment of library instruction relied heavily on student satisfaction surveys. Historically, satisfaction surveys were a primary vehicle for academic support services.<sup>297</sup> Unfortunately, while satisfaction surveys “can yield valuable information, they do not allow the organization to take an intimate look at itself.”<sup>298</sup> In 1981, King wrote that:

The most prevalent of the methodologies currently in use for evaluation of instruction programs is the questionnaire designed to elicit the opinions and perceptions of students and/or faculty concerning the value of instruction, the quality of presentation, the relevance of content, and attitudes toward the library.... These studies may [attempt] to identify user needs and satisfaction. More commonly seen are short questionnaires of the ‘Did you like it?’ variety. Questionnaires often request another sort of opinion as well. Students are commonly asked to assess their own abilities and the extent to which they think they have learned from instruction.<sup>299</sup>

In 1999, Young and Harmony also report the reliance of library instruction assessment on satisfaction surveys:

One of the most popular means of evaluating instruction is the student survey or questionnaire to glean the students’ perceptions. Through questionnaires, you can discover if students “liked” the method used to teach the class and if students “think” they learned something from the presentation. This ‘reaction data’ is the most frequently used method to assess the effectiveness of library instruction.<sup>300</sup>

---

<sup>297</sup> Bresciani, Zelná and Anderson, Assessing Student Learning and Development. 33.

<sup>298</sup> Bresciani, Zelná and Anderson, Assessing Student Learning and Development. 33.

<sup>299</sup> David N. King and John C. Ory, "Effects of Library Instruction on Student Research: A Case Study," College and Research Libraries 42.1 (1981). 32.

<sup>300</sup> Rosemary Young, M. and Stephen Harmony, Working with Faculty to Design Undergraduate Information Literacy Programs, How-to-Do-It Manuals for Librarians (New York: Neal-Schuman, 1999). 76.

DeFranco and Bleiler found that seventy-nine percent of library instruction surveys ask students to assess the instructor's presentation skills.<sup>301</sup> Similarly, Grassian and Kaplowitz find that librarians rely on asking students whether or not they liked the programs and the librarians who taught them.<sup>302</sup> They lament that "rarely do we delve deeper into the questions of whether or not they had actually learned anything."<sup>303</sup> Wolff questions this approach, suggesting that student satisfaction data conveys little about student learning:

What does 'satisfaction' mean? Does it mean satisfaction with the library as a research entity that enabled a user to get a paper done or as a support entity that did some preliminary research? What would happen if the goal of satisfaction surveys was to determine how effectively the library enriched the course or curriculum or educational experience of the student? Or to determine whether library staff had an impact in helping students become lifelong learners?<sup>304</sup>

Wolff goes on to point out that high student satisfaction may or may not connect to actual student learning.<sup>305</sup> He writes, "satisfaction with the hours of operation, library environment, and the friendliness of library staff may not assure that students have learned needed information literacy skills."<sup>306</sup> Indeed, in 2003, fifty-seven percent of surveyed institutions reported that their current assessment tool is "not able to provide adequate information about the success of the

---

<sup>301</sup> Francine DeFranco and Richard Bleiler, Spec Kit #279: Evaluating Library Instruction (Washington, D.C.: Association of Research Libraries, 2003). 15.

<sup>302</sup> Grassian and Kaplowitz, Information Literacy Instruction. 270.

<sup>303</sup> Grassian and Kaplowitz, Information Literacy Instruction. 270.

<sup>304</sup> Wolff, "Using the Accreditation Process." 87-88.

<sup>305</sup> Wolff, "Using the Accreditation Process." 79.

<sup>306</sup> Wolff, "Using the Accreditation Process." 79.

instruction program in the previous year.”<sup>307</sup> Furthermore, they complained that they had difficulty measuring the impact of instruction on student learning.<sup>308</sup> The view that student satisfaction is insufficient to demonstrate student learning is one shared by accreditors.<sup>309</sup>

Three major types of satisfaction surveys, or “affective self-report inventories” exist. They are Likert inventories, Likert-like multidimensional inventories, and confidence inventories.<sup>310</sup> All three are typically administered at the end of a library instruction class,<sup>311</sup> and they are typically voluntary and anonymous.<sup>312</sup>

Likert inventories are typically formatted as a series of statements to which students indicate their level of agreement, usually on a continuum of “strongly agree,” “agree,” “uncertain,” “disagree,” and “strongly disagree.”<sup>313</sup> They are relatively easy to create.<sup>314</sup> The characteristic that distinguishes this type of survey from the next type is that Likert scales are intended to measure only one

---

<sup>307</sup> DeFranco and Bleiler, Spec Kit #279: Evaluating Library Instruction. 15.

<sup>308</sup> DeFranco and Bleiler, Spec Kit #279: Evaluating Library Instruction. 15.

<sup>309</sup> Council for Higher Education Accreditation, “Student Learning Outcomes Workshop,” CHEA Chronicle 5.2 (2002). 2.

<sup>310</sup> Popham, Test Better, Teach Better. 112.

<sup>311</sup> Elizabeth Choiniski, Amy E. Mark and Missy Murphey, “Assessment with Rubrics: An Efficient and Objective Means of Assessing Student Outcomes in an Information Resources Class,” Portal: Libraries and the Academy 3.4 (2003). 564.

<sup>312</sup> DeFranco and Bleiler, Spec Kit #279: Evaluating Library Instruction. 14.

<sup>313</sup> Popham, Test Better, Teach Better. 112.

<sup>314</sup> Popham, Test Better, Teach Better. 113.

affective variable.<sup>315</sup> This can be a concern, since many educators may wish to measure multiple affective variables.<sup>316</sup> Educators seeking to assess multiple variables can use a multidimensional inventory. Multidimensional inventories look like Likert inventories, but they deal with multiple affective variables and measure each variable with two statements, phrased separately.<sup>317</sup> The third type of satisfaction survey is a confidence inventory. A confidence inventory “describes various sorts of activities a student might be asked to engage in, and then it asks students to identify their confidence level as if they were personally required to carry out each activity.”<sup>318</sup> DeFranco and Bleiler found that eighty-six percent of library instruction surveys ask students to assess their own learning.<sup>319</sup> In general, confidence may indicate competence. However, Popham cautions, “Of course there will be obvious exceptions: competent students who lack confidence and incompetent students who ooze confidence.”<sup>320</sup>

Library literature is replete with student satisfaction surveys of all three types, especially multidimensional inventories (although many do not measure each affective variable with at least two items). An excellent collection of such surveys is provided in Evaluating Library Instruction: Sample Questions, Forms,

---

<sup>315</sup> Popham, Test Better, Teach Better. 112-114.

<sup>316</sup> Popham, Test Better, Teach Better. 112-114.

<sup>317</sup> Popham, Test Better, Teach Better. 115.

<sup>318</sup> Popham, Test Better, Teach Better. 117.

<sup>319</sup> DeFranco and Bleiler, Spec Kit #279: Evaluating Library Instruction. 15.

<sup>320</sup> Popham, Test Better, Teach Better. 117.



and Strategies for Practical Use.<sup>321</sup> A more recent collection is available in SPEC Kit #279: Evaluating Library Instruction, published by the Association of Research Libraries.<sup>322</sup>

The long-lasting popularity of satisfaction surveys for library instruction assessment is due in part to the benefits of this approach. Bresciani, Zelna, and Anderson note that the assessment of “student satisfaction, needs, and service utilization is very important. It has great purpose, particularly for [those] who place a heavy emphasis on students’ approval ratings.”<sup>323</sup> Popham believes that the primary reason for educator interest in affective variables (including attitudes, interests, and values) is that they are good predictors of student behavior.<sup>324</sup> He acknowledges the power of “positive attitudes toward learning” and suggests that educators must reinforce positive attitudes and intervene to change negative attitudes.<sup>325</sup> King and Ory demonstrate the commitment of librarians to fostering positive attitudes in students when it comes to information literacy. They state clearly that, “Instilling positive attitudes toward the library and building confidence in library skills are major objectives for most instruction programs.”<sup>326</sup>

In the past, librarians felt satisfaction surveys yielded a number of important results. First, they felt that satisfaction data could help them improve

---

<sup>321</sup> Diana D. Shonrock, Evaluating Library Instruction: Sample Questions, Forms, and Strategies for Practical Use (Chicago: American Library Association, 1996).

<sup>322</sup> DeFranco and Bleiler, Spec Kit #279: Evaluating Library Instruction.

<sup>323</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 19.

<sup>324</sup> Popham, Test Better, Teach Better. 106.

<sup>325</sup> Popham, Test Better, Teach Better. 106-108.

<sup>326</sup> King and Ory, "Effects of Library Instruction." 32.

instruction. Popham writes, "if you discover through affective assessment that your students are really bored or disinterested in the content you're presenting or plan to present, then you can do something to address that problem."<sup>327</sup> Person also has improvement of instruction in mind when he suggests, "If we believe students know a valuable course when they see one, we may conclude that courses in bibliographic instruction are appreciated by those for whom they are designed. We need to document that appreciation, and the reasons for it, in order to improve such instruction."<sup>328</sup> Pre- and post-test comparisons of confidence inventories may also help educators estimate the effectiveness of their instruction.<sup>329</sup> Another possible result of using satisfaction measures is that librarians may be able to document levels of library anxiety among students. Stamatopolos and Mackay suggest that two important goals of library instruction, reducing library anxiety and improving students' confidence in their research skills, can be measured using satisfaction measures, according to Grassian and Kaplowitz.<sup>330</sup>

Two more benefits of satisfaction measures include one political advantage and one practical advantage. Prus and Johnson suggest that satisfaction surveys convey the idea that librarians are interested in the opinions

---

<sup>327</sup> Popham, Test Better, Teach Better. 31.

<sup>328</sup> Roland Person, "Long-Term Evaluation of Bibliographic Instruction: Lasting Encouragement," College and Research Libraries 42.1 (1981). 24.

<sup>329</sup> Popham, Test Better, Teach Better. 120.

<sup>330</sup> Grassian and Kaplowitz, Information Literacy Instruction. 270.

of their constituents.<sup>331</sup> Bresciani, Zelna, and Anderson also point out that there are many works written about survey research and completed surveys to use as models.

Recently, the popularity of student satisfaction surveys has waned because of three main limitations. First, surveys are often poorly designed despite the wealth of research, documentation, and models for this assessment method. Second, the findings of most satisfaction surveys do not provide direct evidence of student learning. Third, the data resulting from satisfaction surveys does not provide educators with materials they need to make improvements to programs and/or instruction.

Poor survey design can greatly undermine any potential benefits of a satisfaction survey approach. Bresciani, Zelna, and Anderson caution that, “designing your own survey is not as easy as it sounds.”<sup>332</sup> They also warn that, “often...surveys are selected as an assessment method prior to articulating any program, student learning, or development outcomes.”<sup>333</sup> Developing a survey before identifying these aspects can result in poor survey design. When poor design is noticed by stakeholders, “it invites criticism and misinterpretation of data and thus poor decision-making.”<sup>334</sup>

Another significant limitation of the satisfaction survey approach to library instruction assessment is that the results often do not appear to help educators

---

<sup>331</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 75.

<sup>332</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development.

<sup>333</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 69.

<sup>334</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 69.

understand to what extent student learning has or has not taken place. Calls for accountability in higher education have set the expectation that all parts of the institution will demonstrate contributions to student learning.<sup>335</sup> While knowing the level of satisfaction of students is valuable, it does not replace knowing whether or not a program is accomplishing its goals regarding student learning and development.<sup>336</sup> Pausch and Popp acknowledge that evaluation tools that measure student attitudes about library instruction do not measure student learning.<sup>337</sup> Choinski, Mark, and Murphy state that not only do satisfaction surveys rarely tell educators what students have learned, they also are not objective.<sup>338</sup> They point out that such surveys are more likely to capture students' satisfaction with the library overall, than specific classes that are taught.<sup>339</sup> Grassian and Kaplowitz warn that these measures "tend to err on the positive side. People are inclined to say they liked the experience because they do not wish to hurt our feelings."<sup>340</sup> Furthermore, Young and Harmony suggest that satisfaction surveys used to assess library instruction capture what students think they've learned, not what they've actually learned.<sup>341</sup> They go on to

---

<sup>335</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 21.

<sup>336</sup> Bresciani, Outcomes Assessment in Student Affairs.

<sup>337</sup> Pausch and Popp, "Assessment of Information Literacy."

<sup>338</sup> Choinski, Mark and Murphey, "Assessment with Rubrics." 564.

<sup>339</sup> Choinski, Mark and Murphey, "Assessment with Rubrics." 564.

<sup>340</sup> Grassian and Kaplowitz, Information Literacy Instruction. 270.

<sup>341</sup> Young and Harmony, Working with Faculty. 77.

recommend that satisfaction surveys not be the only way information literacy programs measure their effectiveness.<sup>342</sup> Grassian and Kaplowitz state:

Asking learners if they feel better is not the same as having them demonstrate this change in some real-life situation. Learners may think they feel better about it all, and may really believe that they have acquired some new skill or ability, but when faced with their next information need, they may still be uncertain, anxious, and unable to apply what they thought they had learned.<sup>343</sup>

Because educators cannot use satisfaction survey data to determine what students have learned, they also cannot use it to make decisions for improving their programs. Grassian and Kaplowitz state that, "Attitude or opinion assessments, though valid in their own right, do not provide all the necessary information to be used for either accountability reporting or a means of improving instruction."<sup>344</sup> Such surveys may indicate areas of strength or weakness, but the level of detail may not be sufficient for decision-making.<sup>345</sup> This is especially true when a program is viewed from the perspective of the entire institution.

Bresciani, Zelna, and Anderson warn:

Findings from [satisfaction survey] assessment[s] do not necessarily help you understand your program's contributions to the greater work of the university. In other words, the assessment findings do not tell you how your program contributes to student development and learning, and the findings seldom help you make decisions for continuous improvement of your programs.<sup>346</sup>

---

<sup>342</sup> Young and Harmony, Working with Faculty. 77.

<sup>343</sup> Grassian and Kaplowitz, Information Literacy Instruction. 270.

<sup>344</sup> Grassian and Kaplowitz, Information Literacy Instruction. 270.

<sup>345</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 72.

<sup>346</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 19.

For example, positive satisfaction results may not actually provide the feedback needed to make future decisions. If a high percentage of students are satisfied with a program, what does that mean? It is difficult or impossible to tell what about the program was useful.<sup>347</sup> Person noted that, “strong satisfaction with [an information literacy] course made it difficult to ascertain any clear weaknesses in the content or method of the course.”<sup>348</sup> On the other hand, negative results are no more useful. “If a program receives feedback registering dissatisfaction, there is not enough information to determine a plan for improvement.”<sup>349</sup> Worse still, when programs invest in the administration and analysis of satisfaction surveys and then have no data with which to make decisions, the commitment necessary for conducting meaning assessments in the future erodes.<sup>350</sup>

As a result of these limitations, many educators have slowly moved from assessing student satisfaction to assessing student learning. This is an evolutionary process. Bresciani, Zelna, and Anderson write, “Most cocurricular professionals do not just decide one day to assess student learning.... They typically try to assess student satisfaction first and then move, over time, to assessing student learning.”<sup>351</sup> Through this transition, educators move from a focus on what educators provide to a focus on what students learn.<sup>352</sup>

---

<sup>347</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 19.

<sup>348</sup> Person, "Long-Term Evaluation of Bibliographic Instruction." 24.

<sup>349</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 19.

<sup>350</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 69.

<sup>351</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 21.

<sup>352</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 21.

### ***Anecdotal Reports***

The third widespread approach to past assessment of library instruction is anecdotal reporting. Barclay lists anecdotal reports as a major method for evaluating library instruction.<sup>353</sup> Warner describes anecdotal reporting as “casual and subjective assumptions about student learning.”<sup>354</sup> She includes feedback from faculty and informal observation of students in class and at the reference desk as the sources for these assumptions.<sup>355</sup> Barclay writes, “Collecting information through anecdotal observation is something librarians do every day. While anecdotal information is important—after all, it is a lifetime of accumulated anecdotal information that makes the experienced librarian valuable—anecdote is subject to personal bias and is often considered soft and unreliable by outsiders (directors, deans, accreditors, etc.).”<sup>356</sup> As a result of these weaknesses, the anecdotal approach falls short in allowing educators to “interpret information with confidence, to discuss the rigor of the research process, and to use feedback in a way that promotes program improvement.”<sup>357</sup>

The past approaches to information literacy instruction assessment—inputs and outputs, satisfaction surveys, and anecdotal reporting—seem to indicate that librarians “make the mistake of designing things around the needs of

---

<sup>353</sup> Barclay, "Evaluating Library Instruction."

<sup>354</sup> Warner, "Programmatic Assessment." 169.

<sup>355</sup> Warner, "Programmatic Assessment." 169.

<sup>356</sup> Barclay, "Evaluating Library Instruction."

<sup>357</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 2.

librarians and not around the needs of students and researchers.”<sup>358</sup> Grassian and Kaplowitz suggest other four reasons for the historical lack of rigor in information literacy assessment. They write, “Assessment takes planning, skill, time and often money. IL instructors develop and deliver instruction under considerable constraints in all those areas. They may feel they lack the expertise to do it properly. In addition, they may be reluctant to siphon off either development time or money for this purpose.”<sup>359</sup> Barclay concludes that the main reasons for poor library instruction assessment is that assessment is too hard and too time consuming. He writes:

As Sugranes and Neal have pointed out, the simple, unsinister reasons teaching librarians tend to avoid evaluation are that evaluation is seen as too complex and too time consuming.... This is especially true for those teaching librarians.... With so much to do, it is not surprising that when the work-a-day...teaching librarian must choose between getting more tasks accomplished—more classes taught, more hours served on the desk, more books selected for the collection—and evaluating a task that has already been completed, evaluation comes last.... Of course, just because teaching librarians are busy people doesn’t mean that the demands for evaluation of library instruction will go away.<sup>360</sup>

Indeed, demands for the assessment of information literacy assessment have grown over the last few years, and many librarians are now struggling to assess in more meaningful and useful ways.

### **The Call for Change in Assessment**

Since the 1980s and 1990s, accreditors have called for a shift from input

---

<sup>358</sup> Lichtenstein, "Informed Instruction." 26.

<sup>359</sup> Grassian and Kaplowitz, Information Literacy Instruction. 269.

<sup>360</sup> Barclay, "Evaluating Library Instruction."



measures to outcome measures.<sup>361</sup> This new model for assessment requires institutions to clarify and publicize their standards for quality, gather and analyze data to see how well the standards are being achieved, and use the results to “document, explain, and improve” performance.<sup>362</sup> As institutions move to this outcomes-based model of assessment, it’s important that libraries follow suit. In fact, Smith argues that the shared need to focus on outcomes, rather than traditional input measures, creates an opportunity for the library to “become an even more central part of the university learning community.”<sup>363</sup> He argues that for libraries to be successful in this new assessment environment, they must understand how universities are adjusting to the new environment, how the new environment impacts the library’s mission, and how the library can become a core part of the university response to the new environment.<sup>364</sup> Smith suggests that the libraries must move from a “content view” to a “competency view” in which the focus is on how the library actively contributes to measurable advances in student learning.<sup>365</sup> He describes this paradigm shift and the resulting focus on student learning and experiences:

The focus on learning involves looking at the academic program not from the perspective of its subject matter content but from the perspective of the competencies to be developed by students.... The focus of outcomes assessment is on the collective success of the program in developing the competencies of the students in the program.... The relevance of learning

---

<sup>361</sup> Smith, New Roles and Responsibilities for the University Library. 3.

<sup>362</sup> Smith, New Roles and Responsibilities for the University Library. 3-4.

<sup>363</sup> Smith, New Roles and Responsibilities for the University Library. 16, 1.

<sup>364</sup> Smith, New Roles and Responsibilities for the University Library. 1.

<sup>365</sup> Smith, New Roles and Responsibilities for the University Library. 9, 17.

as a central concept is that it requires us to focus attention on the student's experience. It requires that we rethink the curriculum, moving from a model in which we package knowledge around the expertise of faculty to a model based on the learning outcomes realized by students. These outcomes include not only what students know, but also the skills they develop, what they are able to do and the attitudes of mind that characterize the way they will approach their work over a lifetime of change.<sup>366</sup>

Finally, Smith calls libraries to view their mission from the students' perspective, to question whether or not the structure of library programs and instruction improves student learning, and to focus attention on a "shared academic culture dedicated to understanding what we are doing and how well we are doing it."<sup>367</sup>

Grassian and Kaplowitz also call for changes in the assessment of library instruction, although they do not refer directly to outcomes-based assessment. Instead, they note academia's new focus on life-long learning skills and acknowledge what they call a "reexamination of how to measure educational competencies."<sup>368</sup> Grassian and Kaplowitz mention new attention paid to the relative merits of qualitative and quantitative assessment methods and suggest that librarians become "more versatile and flexible" in their use of assessment techniques.<sup>369</sup> They suggest that librarians "must be willing to embrace the

---

<sup>366</sup> Smith, New Roles and Responsibilities for the University Library. 6.

<sup>367</sup> Smith, New Roles and Responsibilities for the University Library. 3.

<sup>368</sup> Grassian and Kaplowitz, Information Literacy Instruction. 287.

<sup>369</sup> Grassian and Kaplowitz, Information Literacy Instruction. 287.

notion that different methods are valid under different circumstances. They must be knowledgeable about the strengths and weaknesses of each technique.”<sup>370</sup>

### **Increased Attention in Library Literature**

In recent years, librarians have witnessed a renewed emphasis on assessment in library instruction literature. Many authors call for change in assessment focus and practice. According to Carter’s 2002 review of library literature, the evaluation of library instruction that occurred in the past “was not meaningful, not an integral part of the [library instruction] program, and focused on user satisfaction, not competencies.”<sup>371</sup> In 2003, Warner reviewed library literature and noted her concern about “the absence of a model for programmatic assessment and the absence of examples that used assessment in order to improve teaching.”<sup>372</sup> She writes, “While some worthy case studies had been conducted, no one had solved the problem of moving beyond primarily subjective assessments to a sustained objective assessment of the library instruction program with an interest in programmatic improvement.”<sup>373</sup> In a 2004 publication produced by the Educational Testing Service, authors note:

Despite broad consensus on the need for information and communications technology literacy among college students, until recently there has been a relative dearth of research-based data to isolate exactly what proficiency gaps exist, or what measures to take to ensure that students enter college

---

<sup>370</sup> Grassian and Kaplowitz, Information Literacy Instruction. 287.

<sup>371</sup> Carter, “‘Doing the Best You Can with What You Have’.” 36.

<sup>372</sup> Warner, “Programmatic Assessment.” 170.

<sup>373</sup> Warner, “Programmatic Assessment.” 169.

and the economy prepared for the complexities of an information-driven society.”<sup>374</sup>

The growth of assessment in the literature might be attributable to the increased focus on assessment by practicing academic librarians. In 2002, Merz and Mark found that fifty-nine percent of library instruction programs include assessment.<sup>375</sup> A year later, DeFranco and Bleiler report that two-thirds of libraries assess instruction.<sup>376</sup>

It should be noted that not all of library instruction assessment projects focus on student learning. In 2003, Knight notes that library literature on the evaluation of instruction seems to focus more on the librarian than the student. She summarizes, “In general [the literature] focus[es] on methods for one-time, formative assessment of the librarian, the instructional session, or the overall program. The literature less frequently reports systematic plans for information gathering that measure changes in the students’ skills and confidence as a consequence of library instruction.”<sup>377</sup> Knight concludes that “librarians should continue to develop assessment methods that measure student progress and inform the process of instruction.”<sup>378</sup> Even manuals that seek to guide librarians in conducting assessments of instruction provide examples of this focus on teaching, rather than learning. One such source focuses much more on the

---

<sup>374</sup> Educational Testing Service, ICT Literacy Assessment: An Issue Paper from ETS (2004).

<sup>375</sup> Lawrie H. Merz and Beth L. Mark, Clip Note #32: Assessment in College Library Instruction Programs (Chicago: Association of College and Research Libraries, 2002). 8.

<sup>376</sup> DeFranco and Bleiler, Spec Kit #279: Evaluating Library Instruction. 14.

<sup>377</sup> Knight, “The Role of Assessment in Library User Education.”

<sup>378</sup> Knight, “The Role of Assessment in Library User Education.”

evaluation of the librarian's performance than the student's achievement of learning outcomes, stating:

If the use of the evaluation is for self-improvement of the instructor, highly detailed information that describes the strengths and weaknesses of the instructor is generally preferred (a formative evaluation). If the use is for personnel decisions, then information that measures the overall competence of the instructor is preferred (a summative evaluation).... If the purpose [of evaluating information literacy instruction] is to improve the teaching skills of the librarian, and ideally this is the case, then formative evaluation as opposed to summative evaluation should be used.<sup>379</sup>

The emphasis in this source and others is on teaching, rather than learning. This emphasis must shift if libraries are to contribute to the outcomes-based learning assessments taking place throughout institutions of higher education. Dugan and Hernon note that while many library evaluators continue to focus on the perspective of the library, recent signs show that others are shifting attention to library stakeholders and the contribution of the library to the larger educational mission of the university.<sup>380</sup>

### **Current Methods of Assessing Library Instruction**

Library instruction can be assessed within the library, in the classroom, on campus, and beyond campus.<sup>381</sup> One advantage to conducting assessment within the library is autonomy for librarians. In the library, librarians can conduct, analyze, and report assessment independently.<sup>382</sup> Iannuzzi lists evaluation

---

<sup>379</sup> Young and Harmony, *Working with Faculty*. 74.

<sup>380</sup> Dugan and Hernon, "Outcomes Assessment." 376.

<sup>381</sup> Iannuzzi, "We Are Teaching, but Are They Learning." 304.

<sup>382</sup> Iannuzzi, "We Are Teaching, but Are They Learning." 304-305.

measures that can be used to assess library instruction in the library including in-class assignments or activities, online tutorials, workbook exercises, and tests.<sup>383</sup> Feinberg and King list assessment methods such as pencil and paper tests, performance tests, and "the ubiquitous opinion/attitude questionnaire."<sup>384</sup> Warner reports using a journal tool for students and a faculty reflection tool in her assessment efforts.<sup>385</sup> In Sampson's review of the tools reported in library assessment literature, she includes classroom assessment techniques, focus groups, and portfolio assessment.<sup>386</sup> Flaspohler provides a case study that evaluates student bibliographies to assess library instruction.<sup>387</sup> Of all these methods of assessing library instruction, the most common tools are surveys, case studies, and pre-tests and post-tests administered to students who participated in course-integrated instruction, for-credit courses, or tutorials.<sup>388</sup>

### **Focus on Learning Outcomes**

In 1998, the ACRL Task Force on Academic Library Outcomes recommended that librarians develop qualitative assessment instruments to

---

<sup>383</sup> Iannuzzi, "We Are Teaching, but Are They Learning." 304-305.

<sup>384</sup> Feinberg and King, "Performance Evaluation in Bibliographic Instruction." 75.

<sup>385</sup> Warner, "Programmatic Assessment." 171.

<sup>386</sup> Samson, "What and When Do They Know? Web-Based Assessment."

<sup>387</sup> Molly R. Flaspohler, "Information Literacy Program Assessment: One Small College Takes the Big Plunge," Reference Services Review 31.2 (2003). 135-146.

<sup>388</sup> Research and Scholarship Committee of the ACRL Instruction Section, "Research Agenda." 112.

measure...information literacy standards using outcomes.<sup>389</sup> Learning outcomes are “the essential and enduring knowledge, abilities (skills), and attitudes (values, dispositions) that constitute the integrated learning needed by a graduate of a course or program.”<sup>390</sup> Adopting a learning outcomes approach to instruction and assessment allows educators to make learning more meaningful and more effective.<sup>391</sup> This approach emphasizes the application and integration of knowledge.<sup>392</sup> The leading question of outcomes-focused instruction is “What do students need to know and be able to do after they graduate (from this course, from this program, from the university)?”<sup>393</sup>

In order to adopt a learning outcomes approach to instruction and assessment, it is helpful to define a common language. Among the most important terms to define are “goals”, “objectives”, and “outcomes”. The terms goals and objectives can be used interchangeably to describe the “broad general statements of what a program intends to accomplish.”<sup>394</sup> Goals and objectives are evaluated by measuring the outcomes related to the objective.<sup>395</sup>

---

<sup>389</sup> Iannuzzi, "We Are Teaching, but Are They Learning." 304.

<sup>390</sup> Mark Battersby, "So, What's a Learning Outcome Anyway?," (Vancouver, B.C: Centre for Curriculum Transfer and Technology, 1999).

<sup>391</sup> Battersby, "So, What's a Learning Outcome Anyway?."

<sup>392</sup> Battersby, "So, What's a Learning Outcome Anyway?."

<sup>393</sup> Battersby, "So, What's a Learning Outcome Anyway?."

<sup>394</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 10-11.

<sup>395</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 11.

Outcomes “describe the end result of the program” and must be measurable.<sup>396</sup> In other words, educators must be able to identify or observe student behaviors that show whether or not students know and can do what the program set out to teach them.<sup>397</sup> There are two large general classes of outcomes: (1) program outcomes and (2) student learning and development outcomes. Program outcomes “illustrate what you want your program to accomplish.”<sup>398</sup> Student learning and development outcomes “depict cognitive abilities, as well as affective dimensions that you desire your program to instill or enhance.”<sup>399</sup> These outcomes are not about “what you are going to do to the student, but rather what you want to student to know or do as a result of an initiative, course or activity.”<sup>400</sup> To be measurable, outcomes usually include active verbs that can be identified or observed by assessors.<sup>401</sup> Dugan and Hernon define outcomes in a library context as: “observed, reported, or otherwise quantified changes in attitudes or skills of students on an individual basis because of contact with library services, programs, or instruction.”<sup>402</sup>

---

<sup>396</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 11.

<sup>397</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 11.

<sup>398</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 11.

<sup>399</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 11.

<sup>400</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 11.

<sup>401</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 11.

<sup>402</sup> Dugan and Hernon, "Outcomes Assessment." 379.



Lichtenstein reports that librarians are increasingly asked to join other educators in justifying their programs using learning outcomes.<sup>403</sup> Smith 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."<sup>404</sup> However, Lichtenstein laments that library instruction outcomes are often unstated or too vague.<sup>405</sup> Johnson suggests that library instruction outcomes should focus on higher-order thinking skills and be measured against standards, not against student performance.<sup>406</sup> Lichtenstein suggests that librarians benefit when outcomes are clearly stated, especially as they "serve as the means by which the success of a lesson may be evaluated."<sup>407</sup> Johnson takes the idea one step further by aligning outcomes not just with assessment, but specifically with authentic assessment methods. She writes, "Not surprisingly, outcomes are directly associated with authentic assessment. The underlying theory is that outcomes need to be authentic: that is, true to life and reflecting lifelong learning goals."<sup>408</sup>

---

<sup>403</sup> Lichtenstein, "Informed Instruction." 28.

<sup>404</sup> Smith, New Roles and Responsibilities for the University Library.

<sup>405</sup> Lichtenstein, "Informed Instruction." 28.

<sup>406</sup> Farmer, "Authentic Assessment of Information Literacy."

<sup>407</sup> Lichtenstein, "Informed Instruction." 28.

<sup>408</sup> Farmer, "Authentic Assessment of Information Literacy."

## Outcomes-Based Education

Student learning outcomes focus on “attributes and abilities, both cognitive and affective, which reflect how the student experiences at [an] institution support...their development as individuals.”<sup>409</sup> They include not just what students know, but the “skills they develop, what they are able to do, and the attitudes of mind that characterize the way they will approach their work over a lifetime of change.”<sup>410</sup> In recent years, learning outcomes have become the centerpiece of educational assessment policies and programs. Accreditors, legislators, and other stakeholders now demand that higher education show evidence of quality using student learning outcomes.<sup>411</sup>

As a result of increased attention to student learning outcomes and the growing need to demonstrate effectiveness, institutions of higher education have adopted outcomes-based education principles. Spady describes the process of developing an outcome based education program:

Start by developing a clear picture of what learners should ultimately be able to do successfully at the end of a significant educational experience (i.e., the outcome). Then base (i.e., develop) the curriculum, instruction, assessment, and reporting (i.e., the education) directly on that clear picture. This is a simple matter of clearly defining what one wants learners to be able to do (the end) before the beginning, teaching them how to accomplish that end, and then assessing and documenting the end they were to achieve in the first place. Notice the fundamental cause-and-effect logic of this model: Education (the means) is based on the outcome (the end), not the other way around.<sup>412</sup>

---

<sup>409</sup> Dugan and Hernon, "Outcomes Assessment." 377.

<sup>410</sup> Smith, New Roles and Responsibilities for the University Library.

<sup>411</sup> Iannuzzi, "We Are Teaching, but Are They Learning." 304.

<sup>412</sup> William G. Spady, "Outcome Based Education," Encyclopedia of Education, ed. James W. Guthrie, vol. 5 (New York: Macmillan Reference USA, 2003). 1827-1828.

Outcomes-based education led to the development of standards focused on subject areas in the late 1990s and early 2000s.<sup>413</sup> Outcomes-based education also took what Spady calls another “interdisciplinary and change-oriented” tack.<sup>414</sup> Proponents of this type of outcomes-based education believed that learning, especially learning of higher-order skills, could not and should not be aligned with specific disciplines.<sup>415</sup> They suggested that communication, critical thinking, planning, and problem-solving are examples of skills that require “an elevated, more complex notion of learning and competence; a developmental approach to curriculum design and instruction; and authentic approach to assessment and reporting.”<sup>416</sup> Unfortunately, this interdisciplinary approach to outcomes-based education has been stymied by national and state emphasis on disciplinary-defined outcomes and standardized tests.<sup>417</sup> Still, some efforts continue to pursue an interdisciplinary approach to outcomes-based education. Smith calls librarians to focus on students’ experiences, outside of disciplinary boundaries.<sup>418</sup> He suggests that they “rethink the curriculum, moving from a model in which we package knowledge around the expertise of faculty to a model based on the learning outcomes realized by students.”<sup>419</sup> Smith advocates a

---

<sup>413</sup> Spady, "Outcome Based Education." 1829-1830.

<sup>414</sup> Spady, "Outcome Based Education." 1830.

<sup>415</sup> Spady, "Outcome Based Education." 1830.

<sup>416</sup> Spady, "Outcome Based Education." 1830.

<sup>417</sup> Spady, "Outcome Based Education." 1830.

<sup>418</sup> Smith, New Roles and Responsibilities for the University Library.

<sup>419</sup> Smith, New Roles and Responsibilities for the University Library.

shift in focus from the “teacher’s knowledge to the student’s understandings and capabilities.”<sup>420</sup> This shift—from teacher-centered approach to student-centered approach—requires that both instruction and the assessment of instruction be based on learning outcomes.

### **Outcomes-Based Assessment**

Smith defines assessment as a “means for organizing a conversation among the faculty and other professionals responsible for an academic program.”<sup>421</sup> He identifies five goals of assessment: “to understand...students, determine learning outcomes required for student success, identify how the academic program achieves desired learning outcomes, measure the extent to which outcomes are achieved, and use the knowledge to improve academic programs.”<sup>422</sup> Smith states that the focus of outcomes-based assessment is on the collective success of the program in developing student competencies.<sup>423</sup> Dugan and Hernon point out that outcomes-based assessment goes far beyond measuring student expectations, service quality, and user satisfaction.”<sup>424</sup> Smith summarizes, “The assessment of student outcomes is a means of focusing our collective attention, examining our assumptions and creating a shared academic

---

<sup>420</sup> Smith, New Roles and Responsibilities for the University Library.

<sup>421</sup> Smith, New Roles and Responsibilities for the University Library.

<sup>422</sup> Smith, New Roles and Responsibilities for the University Library.

<sup>423</sup> Smith, New Roles and Responsibilities for the University Library.

<sup>424</sup> Dugan and Hernon, "Outcomes Assessment." 380.

culture dedicated to understanding what we are doing and how well we are doing it and to improving the quality of learning that results.”<sup>425</sup>

Although outcomes-based assessment has been adopted in other parts of institutions of higher education, it remains a challenge for academic libraries.

Iannuzzi reports that libraries find assessment of outcomes-based approaches to instruction a struggle.<sup>426</sup> She states, “We have yet to see widespread

implementation of outcomes assessment methodologies in terms of student learning in our academic libraries.”<sup>427</sup> Iannuzzi offers a few explanations for the challenge of assessing student achievement of information literacy outcomes.

She admits that assessment is “difficult and potentially frightening.”<sup>428</sup> She voices concerns that assessment is too difficult for libraries to complete alone and explains that librarians may be frightened that they will be held solely accountable for any disappointing results.<sup>429</sup> Knight also acknowledges the challenge of assessing student learning in the library.<sup>430</sup>

Despite these concerns and challenges, librarians must fulfill their roles in the assessment of higher education. Blixrud writes, “Learning and its assessment have become a focus of attention at many academic campuses and the role of the library in teaching and learning emerged...as an area in which

---

<sup>425</sup> Smith, New Roles and Responsibilities for the University Library.

<sup>426</sup> Iannuzzi, “We Are Teaching, but Are They Learning.” 304.

<sup>427</sup> Iannuzzi, “We Are Teaching, but Are They Learning.” 304.

<sup>428</sup> Iannuzzi, “We Are Teaching, but Are They Learning.” 304.

<sup>429</sup> Iannuzzi, “We Are Teaching, but Are They Learning.” 304.

<sup>430</sup> Knight, “The Role of Assessment in Library User Education.”

measures are urgently needed.”<sup>431</sup> Librarians must develop assessment methods that measure outcomes of student learning,<sup>432</sup> as these outcomes will “illustrate and demonstrate...the academic library’s value as an institutional teaching and learning partner.”<sup>433</sup> Indeed, outcomes-based assessment offers an important opportunity to show the contribution the library makes to the mission of the institution.<sup>434</sup>

### **Benefits of Outcomes-Based Assessment**

Outcomes-based assessment provides many benefits over other types of assessment. The most important benefit is perhaps that outcomes-based assessment results in improvements to the teaching and learning process.<sup>435</sup>

Carter explains:

Outcomes assessment alerts us to what students know about or do not know about library research, thus allowing librarians to adapt instruction to the needs of the students. It also helps us to determine what we are doing right and what we are doing wrong, what needs more emphasis, and what students already ‘get.’ In short, our instruction is better because we know how we are doing.<sup>436</sup>

There are also program benefits to outcomes-based assessment. Outcomes-based assessment can help librarians clarify the goals of an instruction program,

---

<sup>431</sup> Julia C. Blixrud, Establishing a Role for Research Libraries in Learning Outcomes Assessment, December 2000 2000, Association of Research Libraries <<http://www.arl.org/newsltr/213/assess.html>>.

<sup>432</sup> Iannuzzi, "We Are Teaching, but Are They Learning." 304.

<sup>433</sup> Dugan and Hernon, "Outcomes Assessment." 380.

<sup>434</sup> Dugan and Hernon, "Outcomes Assessment." 378.

<sup>435</sup> Gratch Lindauer, "Comparing the Regional Accreditation Standards." 14.

<sup>436</sup> Carter, "Doing the Best You Can with What You Have'." 41.

examine how they attempt to achieve these goals, and determine the extent to which the goals are achieved.<sup>437</sup> One result of Carter's assessment project was to decrease the traditional reference responsibilities of librarians and increase their instructional responsibilities.<sup>438</sup> She reports, "Through our instruction program we reach faculty and students in an organized and controlled environment, one where we can show results."<sup>439</sup> A third benefit of an outcomes-based approach to assessment is that it offers measurable and meaningful answers to questions of accountability.<sup>440</sup> Gratch Lindauer emphasizes this benefit in the context of the overall institution. She states, "Probably the most direct contribution the library makes to institutional goals is its role in developing clear student learning objectives for information literacy skills; assessing progress and achievement of these objectives; and showing how the outcomes are used to improve student learning."<sup>441</sup>

There are a wide variety of assessment tools that can be used to conduct outcomes-based assessment. Before selecting a tool for outcomes-based assessment, Bresciani, Zelna, and Anderson recommend answering three questions: "Which outcomes do you want to measure?, What do you need to know in order to determine that students know or can do what you have identified

---

<sup>437</sup> Carter, "'Doing the Best You Can with What You Have'." 41.

<sup>438</sup> Carter, "'Doing the Best You Can with What You Have'." 41.

<sup>439</sup> Carter, "'Doing the Best You Can with What You Have'." 41.

<sup>440</sup> Research and Scholarship Committee of the ACRL Instruction Section, "Research Agenda." 112.

<sup>441</sup> Gratch Lindauer, "Comparing the Regional Accreditation Standards." 19.

in the outcome(s)?, Are there set criteria in place or do you need to create the criteria?”<sup>442</sup>

### **Outcomes-Based Surveys and Standardized Tests**

In academic libraries, outcomes-based surveys, unlike satisfaction surveys, “provide information about the students’ library skills before and after a sequence of library instruction and/or research activities.”<sup>443</sup> Some libraries have endeavored to transform outcomes-based surveys into standardized tests in order to administer and score the assessments in a standard, predetermined way<sup>444</sup> and to strive for objectivity.<sup>445</sup> Most outcomes-based surveys and standardized tests focus on multiple-choice or true/false items.<sup>446</sup> Such items and assessments are examples of indirect measures of student learning. Indirect assessments involve “an estimate of the examinee’s probable skill level based on observances of knowledge about skill level (i.e., to assess writing, one would observe vocabulary, grammar, sentence structure, etc.),” as opposed to direct assessments, which might include constructed response items.<sup>447</sup> “Indirect

---

<sup>442</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 25.

<sup>443</sup> Knight, "The Role of Assessment in Library User Education."

<sup>444</sup> Popham, Test Better, Teach Better. 125.

<sup>445</sup> Johanna V. Crighton, "Standardized Tests and Educational Policy," Encyclopedia of Education, ed. James W. Guthrie, vol. 7 (New York: Macmillan Reference USA, 2003). 2530.

<sup>446</sup> Crighton, "Standardized Tests and Educational Policy." 2530.

<sup>447</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment. 4



assessments are exemplified by many standardized commercially available tests.”<sup>448</sup>

Standardized tests are widely regarded as trustworthy measures of student achievement. Popham acknowledges, “In the United States today, most citizens regard students’ performances on standardized achievement tests as the definitive indicator of school quality. These test scores...mark school staffs as either successful or unsuccessful.”<sup>449</sup> Because the public believes that standardized tests can measure what students learn in school, legislators and some administrators believe that student scores on these tests “provide a defensible indication of a school’s instructional quality.”<sup>450</sup> What is not often acknowledged is that the strengths of outcomes-based surveys and standardized tests are also their weaknesses. Standardized tests are designed to measure a student’s ability to answer test items relative to previous test-takers.<sup>451</sup> This ability to compare students, often called “norm-referencing,” is “the cornerstone of standardized achievement testing and has been since standardized testing’s origins in the early 20<sup>th</sup> century.”<sup>452</sup> Without referencing a student’s scores against other test-takers, the test score has no meaning. Popham states, “Raw

---

<sup>448</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment. 4.

<sup>449</sup> Popham, Test Better, Teach Better. 123.

<sup>450</sup> Popham, Test Better, Teach Better. 136.

<sup>451</sup> Popham, Test Better, Teach Better. 125.

<sup>452</sup> Popham, Test Better, Teach Better. 126.

test scores all by themselves are really quite uninterpretable.”<sup>453</sup> Popham asserts that most people who support standardized tests don’t know that “the historic mission of standardized testing is at cross-purposes with the intent of achievement testing. And because of the historic need to produce score spread, standardized achievement tests don’t do a very good job of measuring what students have learned.”<sup>454</sup> Crighton states clearly, “This type of standardization is no longer considered capable of capturing the full range of skills candidates may possess.”<sup>455</sup>

### ***Theoretical Background***

Outcomes-based surveys and standardized tests are assessment measures derived from early models and theories about learning and measurement. According to Shepard, “those measurement perspectives, now felt to be incompatible with instruction, came from an earlier, highly theoretical framework...in which conceptions of ‘scientific measurement’ were closely aligned with traditional curricula and beliefs about learning.”<sup>456</sup>

A number of theorists from the early 20<sup>th</sup> century can be credited with the groundwork for outcomes-based surveys and standardized tests. Beginning with the social efficiency movement of the early 1900s, principles of scientific

---

<sup>453</sup> Popham, Test Better, Teach Better. 126.

<sup>454</sup> Popham, Test Better, Teach Better. 136.

<sup>455</sup> Crighton, "Standardized Tests and Educational Policy." 2530.

<sup>456</sup> Shepard, "The Role of Assessment in a Learning Culture." 4.

measurement were applied to schools.<sup>457</sup> It was believed that every task worth learning could be broken down into fundamental building blocks, which instructors would teach, students would learn, and instructors would measure.<sup>458</sup> To support this model, “precise standards of measurement were required to ensure that each skill was mastered at the desired level.”<sup>459</sup> Edward Thorndike is credited with being the ‘father’ of scientific measurement and originating this associationist learning theory.<sup>460</sup> It was Thorndike who brought these connections together to “foster the development and dominance of the ‘objective’ test, which has been the single most striking feature of academic testing in the United States from the beginning of the century to the present day.”<sup>461</sup> In addition to Thorndike’s associationism, Hull, Skinner, and Gagne contributed behaviorist theories that “conceived of learning as the acquisition of stimulus-response associations.”<sup>462</sup> These stimulus-response theories impacted how student motivation and cognitive development were understood. For example, B.F. Skinner wrote:

The whole process of becoming competent in any field must be divided into a very large number of very small steps, and reinforcement must be contingent upon the accomplishment of each step....By making each successive step as small as possible, the frequency of reinforcement can

---

<sup>457</sup> Shepard, "The Role of Assessment in a Learning Culture." 4.

<sup>458</sup> Shepard, "The Role of Assessment in a Learning Culture." 4.

<sup>459</sup> Shepard, "The Role of Assessment in a Learning Culture." 4.

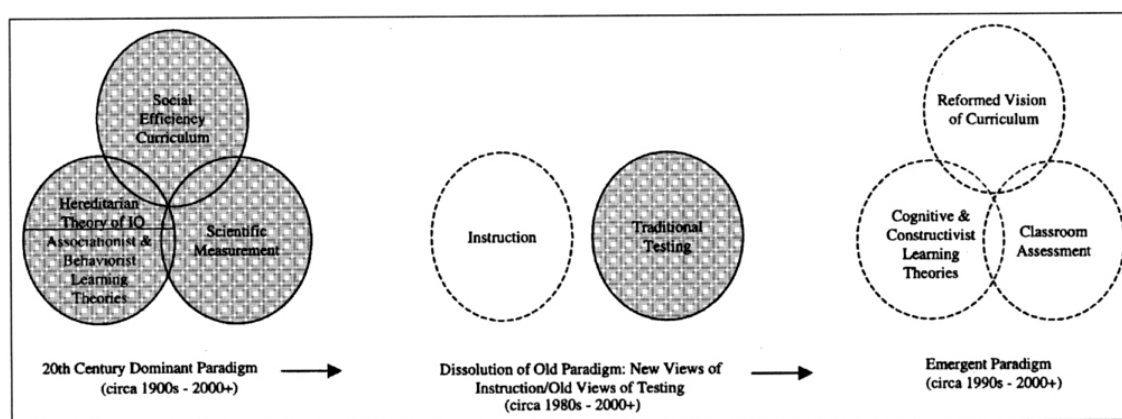
<sup>460</sup> Shepard, "The Role of Assessment in a Learning Culture." 5.

<sup>461</sup> Shepard, "The Role of Assessment in a Learning Culture." 5.

<sup>462</sup> Shepard, "The Role of Assessment in a Learning Culture." 5.

be raised to a maximum, while the possible averse consequences of being wrong are reduced to a minimum.<sup>463</sup>

According to Shepard, recognizing the “common paternity of behaviorist learning theory and objective testing helps us to understand the continued intellectual kinship between one-skill-at-a-time test items and instructional practices aimed at mastery of constituent elements.”<sup>464</sup> To this end, Shepard depicts paradigms that dominate 20<sup>th</sup> century educational measurement (see Figure 3.1). Shepard describes the close links between social efficiency and scientific measurement



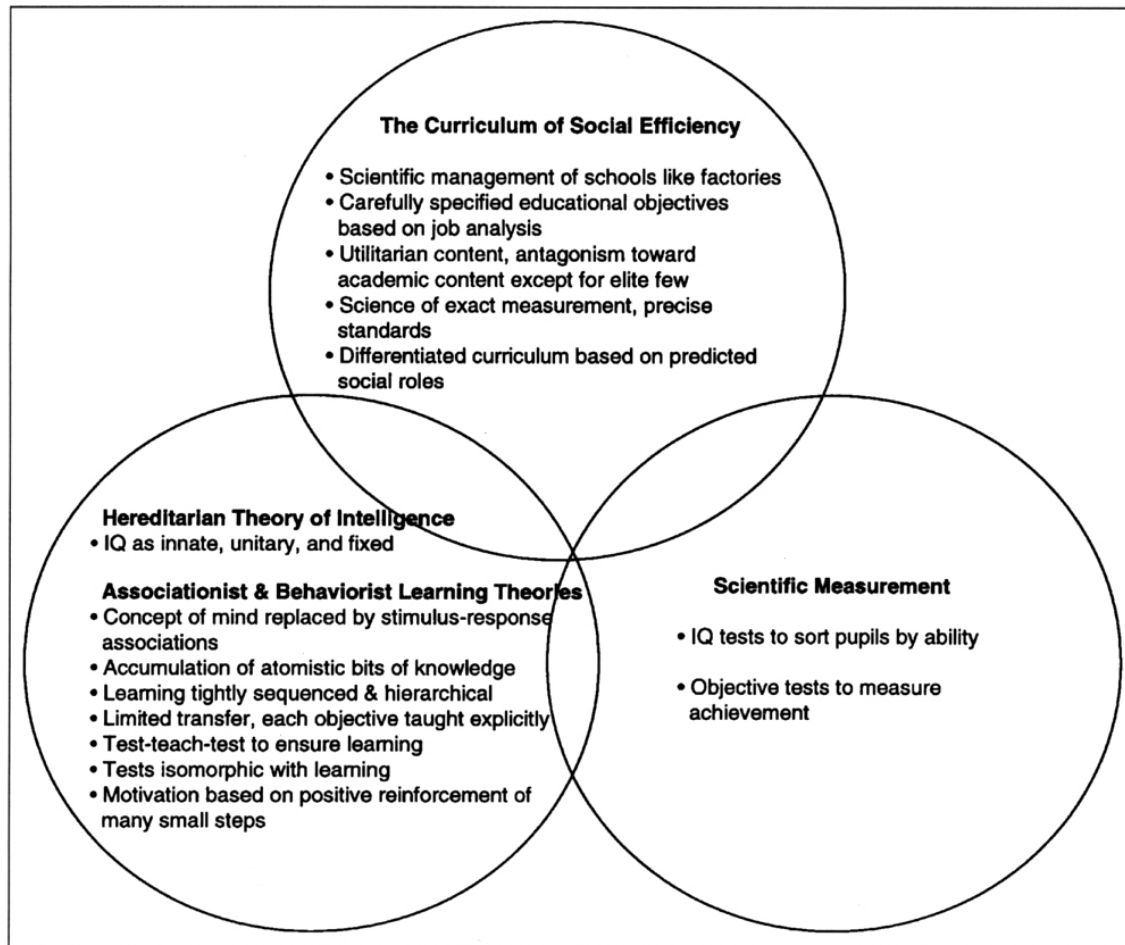
*Figure 3.1. Educational Paradigms*

theories and hereditarian theories of individual differences and associationist and behaviorist learning theories (see Figure 3.2).<sup>465</sup>

<sup>463</sup> Skinner qtd. in Shepard, "The Role of Assessment in a Learning Culture." 5.

<sup>464</sup> Shepard, "The Role of Assessment in a Learning Culture." 5.

<sup>465</sup> Shepard, "The Role of Assessment in a Learning Culture." 4.



*Figure 3.2. Educational Theories*

Early models and theories of learning and measurement have had long-lasting impact on educational assessment methods. For example, Shepard lists six key assumptions from the behaviorist model that affected beliefs about teaching and testing:

1. Learning occurs by accumulating atomized bits of knowledge.
2. Learning is tightly sequenced and hierarchical.
3. Transfer is limited, so each objective must be explicitly taught.
4. Tests should be used frequently to ensure mastery before proceeding to the next objective.
5. Tests are isomorphic with learning (tests = learning).
6. Motivation is external and based on positive reinforcement of many small steps.<sup>466</sup>

<sup>466</sup> Shepard, "The Role of Assessment in a Learning Culture." 5.

Shepard also notes that the emphasis on “rote recall” evident in early tests is still present, despite the changes in what educators know about learning and assessment:

One hundred years ago, various recall, completion, matching, and multiple-choice test types, along with some essay questions, fit closely with what was deemed important to learn. However, once curriculum became encapsulated and represented by these types of items, it is reasonable to say that these formats locked in a particular and outdated conception of subject matter.<sup>467</sup>

These theories continue to impact current assessment practices. Shepard states that past theories operate as the “default framework affecting and driving current practices and perspectives. Belief systems of teachers, parents, and policy makers derive from these old theories.”<sup>468</sup> The old theories not only impact beliefs about how students learn and ways to assess their learning, they also influence beliefs about fairness in testing, the need for separation of testing from teaching, the importance of uniform administration, and the nature of “objectivity.”<sup>469</sup> Shepard emphasizes, “Any attempt to change the form and purpose of classroom assessment to make it more fundamentally a part of the [teaching and learning] process must acknowledge the power of these enduring and hidden beliefs.”<sup>470</sup>

---

<sup>467</sup> Shepard, “The Role of Assessment in a Learning Culture.” 5.

<sup>468</sup> Shepard, “The Role of Assessment in a Learning Culture.” 4.

<sup>469</sup> Shepard, “The Role of Assessment in a Learning Culture.” 5-6.

<sup>470</sup> Shepard, “The Role of Assessment in a Learning Culture.” 6.

## **Benefits**

Despite their deep roots in outdated educational theory, outcomes-based surveys and standardized tests offer a number of benefits. Indeed, librarians have “realized some success with summative assessment devices, such as tests and surveys.”<sup>471</sup> As quantitative measures, surveys and tests provide data in numerical form and are excellent choices to find answers to questions of how much or how many.<sup>472</sup> They are easy to score and require less time and money, especially if computers are used for scoring.<sup>473</sup> In this way, they allow for the collection of a lot of data quickly.<sup>474</sup> Surveys and tests are good tools for measuring students’ acquisition of facts<sup>475</sup> and can be used to compare pre- and post-test results<sup>476</sup> or to compare groups of students to each other.<sup>477</sup>

Another advantage of surveys and tests, especially those made up of multiple-choice items, is that they can be made highly reliable.<sup>478</sup> In fact, high reliability is one of the most frequently cited advantages of surveys and tests.<sup>479</sup>

---

<sup>471</sup> Knight, "The Role of Assessment in Library User Education."

<sup>472</sup> Grassian and Kaplowitz, Information Literacy Instruction. 276.

<sup>473</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment. 4.

<sup>474</sup> Grassian and Kaplowitz, Information Literacy Instruction. 281.

<sup>475</sup> Grassian and Kaplowitz, Information Literacy Instruction. 281.

<sup>476</sup> Knight, "The Role of Assessment in Library User Education."

<sup>477</sup> Grassian and Kaplowitz, Information Literacy Instruction. 281.

<sup>478</sup> Colton, Gao, Harris, Kolen, Martinovich-Barhite, Wang and Welch, Reliability Issues with Performance Assessments. 3.

<sup>479</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment. 4.

One way to increase the reliability of a multiple-choice survey or test is to make it longer.<sup>480</sup> Lengthening a survey or test is much easier than lengthening other types of assessment methods.<sup>481</sup> Furthermore, test/retest and parallel forms reliability estimates are easier to obtain with surveys and tests than other assessment methods that take more time or are difficult to repeat exactly.<sup>482</sup> Indirect assessments like surveys and tests also tend to have a higher predictive validity with “a variety of outcome measures, such as college GPA or scores on other standardized tests.”<sup>483</sup>

Still another advantage to using outcomes-based surveys and tests for assessment is that people believe in them.<sup>484</sup> Because the public is familiar with commercially designed tests and believes them to be extensively developed, tests can be used for “enhanced political leverage.”<sup>485</sup> Policy makers may prefer standardized tests because they compare students’ achievement against other groups or national profiles.<sup>486</sup> Parents and students might also value such

---

<sup>480</sup> Colton, Gao, Harris, Kolen, Martinovich-Barhite, Wang and Welch, Reliability Issues with Performance Assessments. 3.

<sup>481</sup> Colton, Gao, Harris, Kolen, Martinovich-Barhite, Wang and Welch, Reliability Issues with Performance Assessments. 3

<sup>482</sup> Colton, Gao, Harris, Kolen, Martinovich-Barhite, Wang and Welch, Reliability Issues with Performance Assessments. 4.

<sup>483</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment. 4.

<sup>484</sup> Lorrie A. Shepard, "Why We Need Better Assessments," A Handbook for Student Performance Assessment in an Era of Restructuring, eds. Robert E. Blum and Judith A. Arter (Alexandria, Virginia: Association for Supervision and Curriculum Development, 1996). 1-2: 3.

<sup>485</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment. 4.

<sup>486</sup> Maki, "From Standardized Tests to Alternative Methods."



normative comparisons and try to use them to identify an individual student's strengths and weaknesses.<sup>487</sup>

Locally developed outcomes-based surveys have several additional benefits. First, they have the benefit of being adapted to local goals and student characteristics.<sup>488</sup> The process of developing the surveys can help staff determine what they really want to know about student learning.<sup>489</sup> Local grading is an additional benefit—staff have control over the interpretation and use of the results and students receive immediate feedback.<sup>490</sup> Finally, commercially developed standardized tests offer two additional benefits: they can be implemented quickly and they reduce the staff time that would be otherwise used to develop and grade another assessment measure.<sup>491</sup>

### ***Limitations***

“Twenty years ago, standardized tests served as reasonable indicators of student learning. In today’s political climate, tests are inadequate and misleading as measures of achievement.”<sup>492</sup> While outcomes-based surveys and standardized tests are widely used and, in many arenas, still widely respected, most educators now recognize that they “have always been fallible, limited

---

<sup>487</sup> Popham, Test Better, Teach Better. 126.

<sup>488</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 72.

<sup>489</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 71.

<sup>490</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 72.

<sup>491</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 71.

<sup>492</sup> Shepard, "Why We Need Better Assessments." I-2: 4.

measures of learning goals.”<sup>493</sup> Overall, outcomes-based surveys and standardized tests have several limitations: they do not test higher-level thinking skills, they lack authenticity, they tend to have the problems associated with all high-stakes testing, and they are time consuming to create, difficult to analyze, and problematic on a local level.

A major limitation of outcomes-based surveys and standardized tests is that they are indirect assessments that fail to measure higher-order thinking skills.<sup>494</sup> As “objective” tests, they measure low-level recognition rather than recall.<sup>495</sup> Because of artificial time limits and the pressure to survey as much content as possible, outcomes-based surveys and tests rarely involve the interrelation of dimensions of the same topic.<sup>496</sup> By focusing only on individual parts of a concept, test creators tend to develop over-simplified test items. Furthermore, the fixed answer choices provided to students limit the ability of outcomes-based surveys and tests to measure changes in “complex behavior or actual performance success.”<sup>497</sup> Because of these limitations, it is very difficult to use outcomes-based surveys and standardized tests to quantitatively measure the results of improved information literacy instruction.<sup>498</sup> Indirect assessments, like outcomes-based surveys and standardized tests, may “dramatically under-

---

<sup>493</sup> Shepard, "Why We Need Better Assessments." I-2:6.

<sup>494</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment. 4.

<sup>495</sup> Grassian and Kaplowitz, Information Literacy Instruction. 281.

<sup>496</sup> Shepard, "Why We Need Better Assessments." I-2: 3.

<sup>497</sup> Grassian and Kaplowitz, Information Literacy Instruction. 281.

<sup>498</sup> Flaspohler, "Information Literacy Program Assessment." 139.

represent” constructs like writing, critical thinking, and information literacy.<sup>499</sup>

Also, outcomes-based surveys and standardized tests do not necessarily help students learn and develop complex skills, which should ultimately be a goal of good assessment.<sup>500</sup> Grassian and Kaplowitz state that such tests are “less valid for testing higher-level cognitive skills such as analysis, synthesis, and evaluation, or to determine process learning and the acquisition of concepts. As such, they may not be appropriate for many of our needs.”<sup>501</sup>

A second limitation of outcomes-based surveys and standardized tests is their inability to provide an authentic assessment of student learning. These types of tests are “frequently criticized for setting up an artificial situation that does not really test how the learner would react in a real-world situation.”<sup>502</sup> As a result, they sacrifice authenticity “since they differ markedly from the ways in which people apply knowledge in the world outside of school.”<sup>503</sup> One problem is that the conditions of these tests are highly controlled. Students must work within time limits, with limited access to resources, and have few opportunities to make revisions.<sup>504</sup> Because of this, outcomes-based surveys and standardized tests tend to “overassess student ‘knowledge’ and underassess student ‘know-

---

<sup>499</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment. 4.

<sup>500</sup> King and Ory, "Effects of Library Instruction." 145.

<sup>501</sup> Grassian and Kaplowitz, Information Literacy Instruction. 281.

<sup>502</sup> Grassian and Kaplowitz, Information Literacy Instruction. 281.

<sup>503</sup> Robert J. Marzano, Debra Pickering and Jay McTighe, Assessing Student Outcomes: Performance Assessment Using the Dimensions of Learning Model (Alexandria, Virginia: Association for Supervision and Curriculum Development, 1993). 11.

<sup>504</sup> Marzano, Pickering and McTighe, Assessing Student Outcomes. 11.

how with knowledge.”<sup>505</sup> As a result, students who score well on outcomes-based surveys and standardized tests may only be demonstrating that they are good test takers.<sup>506</sup> When faced with a real-world scenario, these students may not be able to formulate an appropriate response.<sup>507</sup> This is a dangerous limitation of outcomes-based surveys and standardized tests, because it may signal to students that the point of learning is not to “acquire ‘useable knowledge,’ but rather ‘testable knowledge.’”<sup>508</sup>

A third important limitation of outcomes-based surveys and standardized tests is that most are designed to produce variance of scores, or “score-spread.” Most standardized tests are intended to allow comparisons among students or groups of students. To do this, the tests must spread out student scores, rather than allow them to bunch together. According to Popham, “It is this quest for score spread that renders such tests unsuitable for the evaluation of school and teacher quality.”<sup>509</sup> He points out that tests designed to compare students are misused when they are misapplied to assess program or institutional effectiveness.<sup>510</sup> Furthermore, the time constraints of most standardized tests exacerbate the problem of score-spread. Time constraints are necessary so that

---

<sup>505</sup> Grant Wiggins, “Creating Tests Worth Taking,” A Handbook for Student Performance in an Era of Restructuring, eds. Robert E. Blum and Judith A. Arter (Alexandria, Virginia: Association for Supervision and Curriculum Development, 1996). V-6: 3.

<sup>506</sup> Grassian and Kaplowitz, Information Literacy Instruction. 277.

<sup>507</sup> Grassian and Kaplowitz, Information Literacy Instruction. 277.

<sup>508</sup> Battersby, “So, What’s a Learning Outcome Anyway?.”

<sup>509</sup> Popham, Test Better, Teach Better. 126.

<sup>510</sup> Popham, Test Better, Teach Better. 127.

students do not become “restless or, worse, openly rebellious,” but they cause test designers to strive for maximum score-spread in the fewest number of test items.<sup>511</sup> As a result, test designers include a large number of items that only forty to sixty percent of the students will answer correctly, as these are the questions that will produce the most score-spread.<sup>512</sup> This means that very few of the questions that eighty to ninety percent of the students will answer correctly are left. As a result, these tests cannot detect effective instruction.<sup>513</sup> Popham explains:

Items with high p-values [high percentage of students answering correctly] indicate that most students possess the knowledge or have mastered the skills that the items represent. The skills and knowledge that teachers regard as most important tend to be the ones that those teachers stress in their instruction. And, even allowing for plenty of differences in teachers’ instructional skills, the more that teachers stress certain content, the better their students will perform on items that measure the teacher-stressed content. But the better students perform on those items, the more likely it will be that those very items will be jettisoned when the standardized test is revised. In short, the quest for score spread creates a clearly identifiable tendency to remove from traditionally constructed standardized achievement tests those items that measure the most important, teacher-stressed content. Clearly, a test that deliberately dodges the most important things teacher try to teach should not be used to judge teachers’ instructional success.<sup>514</sup>

Popham estimates that on tests in subject areas other than math, fifty to eighty-five percent of standardized test questions are designed to spread student scores.<sup>515</sup>

---

<sup>511</sup> Popham, Test Better, Teach Better. 127.

<sup>512</sup> Popham, Test Better, Teach Better. 127.

<sup>513</sup> Popham, Test Better, Teach Better. 128.

<sup>514</sup> Popham, Test Better, Teach Better. 128.

<sup>515</sup> Popham, Test Better, Teach Better. 132.

A fourth significant limitation of outcomes-based surveys and standardized tests of information literacy instruction is that they share characteristics of “high-stakes” tests, and in some cases, they are actually used as high-stakes tests.

Shepard explains:

High-stakes testing is a term that was first used in the 1980s to describe testing programs that have serious consequences for students or educators. Tests are high-stakes if their outcomes determine such important things as promotion to the next grade, graduation, merit pay for teachers, or school rankings reported in a newspaper. When test results have serious consequences, the requirements for evidence of test validity are correspondingly higher.<sup>516</sup>

Because of the high stakes of such assessments, tests that fall into this category must “meet the most stringent technical standards because of the harm to individuals that would be caused by test inaccuracies.”<sup>517</sup> This is problematic because group administered multiple-choice tests “always include a potentially higher degree of error, usually not correctable by ‘guessing correction’ formulae,” resulting in lower test validity.<sup>518</sup>

High-stakes assessments are typically used to monitor and compare groups of students. In aggregate form, educators and administrators can use high-stakes assessment data for decision-making.<sup>519</sup> However, such tests are often inappropriate for measuring individual student performance. Shepard explains, “Because there is not a single national or international curriculum, assessment content must be comprehensive or inclusive of all the curricular

---

<sup>516</sup> Shepard, “Standardized Tests and High-Stakes Assessment.” 2533-2534.

<sup>517</sup> Shepard, “Standardized Tests and High-Stakes Assessment.” 2535.

<sup>518</sup> Prus and Johnson, “A Critical Review of Student Assessment Options.” 71-72.

<sup>519</sup> Shepard, “Standardized Tests and High-Stakes Assessment.” 2534-2535.

goals of the many participating states or nations. Obviously, no one student could be expected to master all of the content in a test spanning many curricula, but by design, individual student scores are not reported in this type of assessment.”<sup>520</sup> Problems can occur when these types of assessments are misused and scores are reported for individual students. Problems can also occur if high-stakes standardized tests are misused in an effort to achieve local goals. For example, such tests “eliminate the important process of learning and clarification of goals and objectives typically associated with local development of measurement instruments.”<sup>521</sup> They also are “unlikely to measure the specific goals and objectives of a program, department, or institution.”<sup>522</sup>

For these reasons, high-stakes tests should not be used as sole determiners of major educational decisions for individual students. Indeed, the Standards for Educational and Psychological Testing state that, “Individual students should be protected from tests being used as the sole criterion for critically important decisions.”<sup>523</sup> Nitko outlines three inappropriate uses of standardized tests. He states that students should not be placed in a special instructional program or retained in a grade solely only results from a standardized test.<sup>524</sup> He also warns that school programs should not be judged

---

<sup>520</sup> Shepard, "Standardized Tests and High-Stakes Assessment." 2534-2535.

<sup>521</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 71-72.

<sup>522</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 71-72.

<sup>523</sup> Shepard, "Standardized Tests and High-Stakes Assessment." 2536.

<sup>524</sup> Anthony J. Nitko, Educational Assessment of Students (Upper Saddle River, New Jersey: Pearson Education, 2004). 376.

solely on the results of a standardized test.<sup>525</sup> Prus and Johnson agree, and caution that standardized test results are “highly susceptible to misinterpretation and misuse both within and outside the institution.”<sup>526</sup> They also state that the results of standardized tests are “unlikely to have direct implications for program improvement or individual student progress.”<sup>527</sup> Popham states clearly, “the primary use of standardized...tests today—to evaluate school and teacher quality—is a misuse.”<sup>528</sup> He writes:

These days, many teachers’ instructional competence is being determined on the basis of a single achievement test...but this sort of teacher-appraisal flunks on several counts. For one thing, it relies on the wrong kind of measurement tool.... The problem is that each year’s testing takes place with a different group of students, and the results depend on the collection of kids being compared.”<sup>529</sup>

Not only do these sorts of tests fail to accurately assess teacher effectiveness and student learning, Shepard argues that they lead to “the de-skilling and de-professionalization of teachers, even...to the denigration of teaching.”<sup>530</sup>

High-stakes testing has many negative implications for teaching and learning. According to Shepard, researchers have determined that testing strongly influences instruction.<sup>531</sup> Specifically, Shepard suggests that high-

---

<sup>525</sup> Nitko, Educational Assessment of Students. 376.

<sup>526</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 71-72.

<sup>527</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 71-72.

<sup>528</sup> Popham, Test Better, Teach Better. 125.

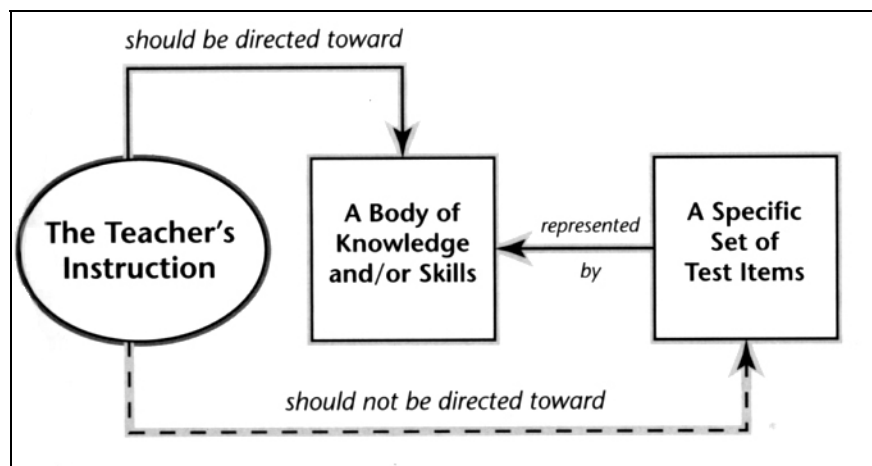
<sup>529</sup> Popham, Test Better, Teach Better. 13.

<sup>530</sup> Shepard, "The Role of Assessment in a Learning Culture." 9.

<sup>531</sup> Shepard, "Why We Need Better Assessments." I-2: 3.



stakes tests lead to score inflation and curriculum distortion.<sup>532</sup> Shepard writes, “Under intense political pressure, test scores are likely to go up without a corresponding improvement in student learning. In fact, distortions in what and how students are taught may actually decrease students’ conceptual understanding.”<sup>533</sup> She notes that teachers will “teach to the test” when student scores have serious consequences.<sup>534</sup> This “cheapens instruction and undermines the authenticity of scores as measures of what children really know.”<sup>535</sup> Popham depicts the need for teachers to teach to knowledge rather than tests (see Figure 3.3).<sup>536</sup>



*Figure 3.3. Teaching for Knowledge, Not the Test*

“Curricular reductionism, wherein teachers have chosen (or have been directed) to give short shrift to any content not assessed on...tests” is another

<sup>532</sup> Shepard, "Standardized Tests and High-Stakes Assessment." 2536.

<sup>533</sup> Shepard, "The Role of Assessment in a Learning Culture." 9.

<sup>534</sup> Shepard, "Why We Need Better Assessments." I-2:3.

<sup>535</sup> Shepard, "Why We Need Better Assessments." I-2: 3.

<sup>536</sup> Popham, Test Better, Teach Better. 135.

major disadvantage of high-stakes testing.<sup>537</sup> Shepard states that it is important to recognize the “pervasive negative effects of accountability tests and the extent to which externally imposed testing programs prevent and drive out thoughtful classroom practices.”<sup>538</sup> For example, standardized tests that are used by large numbers of students and have time limits often rely on multiple-choice item formats.<sup>539</sup> Multiple-choice questions cannot measure a student’s ability to organize information or present arguments.<sup>540</sup> Test development procedures eliminate “the most imaginative and challenging problem-solving tasks”<sup>541</sup> and leave only items that cover “relatively superficial knowledge or learning.”<sup>542</sup> Such test items “force students to perform in one standard, uniform manner. They reinforce the teacher-centric approach to learning that requires rote memorization and exact repetition of what has been presented.”<sup>543</sup> Shepard writes:

Conceiving instruction in the format of multiple-choice items has other far reaching negative consequences: it leads to endless drill and practice on decontextualized skills. The notion that learning comes about by the accretion of little bits is outmoded learning theory. Current models...contend that learners gain understanding when they construct their own knowledge and develop their own cognitive maps of the interconnections among concepts and facts. Thus, real learning cannot be spoon-fed, one skill at a time.<sup>544</sup>

---

<sup>537</sup> Popham, Test Better, Teach Better. 124.

<sup>538</sup> Shepard, "The Role of Assessment in a Learning Culture." 9.

<sup>539</sup> Shepard, "Why We Need Better Assessments." I-2: 3.

<sup>540</sup> Shepard, "Why We Need Better Assessments." I-2: 3.

<sup>541</sup> Shepard, "Why We Need Better Assessments." I-2: 3.

<sup>542</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 71-72.

<sup>543</sup> Grassian and Kaplowitz, Information Literacy Instruction. 278.

<sup>544</sup> Shepard, "Why We Need Better Assessments." I-2: 3.

The item types found on most standardized tests result in the emphasis on basic skills, “limiting the ‘height’ as well as the depth and breadth of permissible content. Even the advocates of high-stakes testing acknowledge that the tests do not cover the full range of important instructional objectives.”<sup>545</sup>

Additionally, Choinski, Mark, and Murphey state that a significant drawback to high-stakes tests is “the need to develop a test instrument separate from what is used for determining a course grade.”<sup>546</sup> Shepard warns that such tests teach students that their efforts should focus on external rewards or punishments and teach teachers to “comply or get out.”<sup>547</sup>

Three more limitations round out the problems associated with high-stakes outcomes-based surveys and standardized tests. Surveys and tests are time-consuming, demand significant resources, and provide inferential results.<sup>548</sup> Grassian and Kaplowitz state that “fixed choice assessments (multiple-choice, true/false, matching) are high on control but are difficult and time consuming to construct. They require a good deal of specialized training to develop and analyze.”<sup>549</sup> Bresciani, Zelna, and Anderson caution: “Sometimes scoring and interpreting standardized tests that assess certain student learning development attributes is not easy. Be sure to find out as much as you are able about the

---

<sup>545</sup> Shepard, “Why We Need Better Assessments.” I-2: 3.

<sup>546</sup> Choinski, Mark and Murphey, “Assessment with Rubrics.” 564.

<sup>547</sup> Shepard, “The Role of Assessment in a Learning Culture.” 9.

<sup>548</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment. 4.

<sup>549</sup> Grassian and Kaplowitz, Information Literacy Instruction. 281.

instrument and also find out if the provider of the instrument can analyze that data for you or offer some assistance with interpretation.”<sup>550</sup> If test data is not analyzed or interpreted, such assessments drain resources. Even so, Ewell and Jones (1993) suggest that “conclusions drawn from indirect indicators [such as surveys and standardized tests] are highly inferential even when the data are presented from multiple measures.”<sup>551</sup>

When outcomes-based surveys and standardized tests are locally developed, a few additional limitations apply. The process of constructing a survey is a difficult one. Locally developed surveys and tests require leadership, coordination, and expertise in measurement.<sup>552</sup> A survey with good psychometric properties can take years to develop,<sup>553</sup> and an outcomes-based survey or test may take even longer since they require not just content expertise, but also expertise in the study of learning.<sup>554</sup> As a result, locally-developed surveys and tests may not provide for “externality” or a “degree of objectivity associated with review and comparisons external to the program or institution.”<sup>555</sup> Even if a local assessment is determined to be adequate, it’s important not to administer the same test so often that students become “over-surveyed”.<sup>556</sup>

---

<sup>550</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 70.

<sup>551</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment. 4.

<sup>552</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 72-73.

<sup>553</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 70.

<sup>554</sup> Smith, New Roles and Responsibilities for the University Library.

<sup>555</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 72-73.

<sup>556</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 71.

### ***Multiple-Choice Items***

Most outcomes-based surveys and standardized tests are made up of selected-response item types, such as multiple-choice, true/false, or matching. Williams cites several reasons for their popularity. For example, selected-response item types can be administered quickly to individuals or groups, allow assessment of factual knowledge, sample a broad range of material in a short time, and do not require a special set up.<sup>557</sup> However, there are a number of tradeoffs to these advantages. Selected-response items usually deal with “low-level, readily memorizable content.”<sup>558</sup> They cannot be used to measure learning outcomes of high-order thinking skills, penalize students who do not read well, are susceptible to guessing, and emphasize getting the “right answer”.<sup>559</sup> They are also “quite narrow in their focus. They provide only a snapshot or a ‘one moment in time’ picture of learning. Although [they] may have certain uses, [they are] generally incapable of revealing in any comprehensive way what students know and can do.”<sup>560</sup>

Many outcomes-based surveys and standardized tests are made up exclusively of multiple-choice questions. Popham asserts, “Rarely will one kind of selected-response test item do such a thorough job that a teacher can come

---

<sup>557</sup> Janet L. Williams, "Creativity in Assessment of Library Instruction," Reference Services Review 28.4 (2000). 333.

<sup>558</sup> Popham, Test Better, Teach Better. 63-64.

<sup>559</sup> Williams, "Creativity in Assessment." 333.

<sup>560</sup> Marzano, Pickering and McTighe, Assessing Student Outcomes. 11.

up with a valid inference about the student's status based on that one item-type alone."<sup>561</sup> Yet, these types of assessments persist. In fact, they are probably the most popular type of selected-response item.<sup>562</sup> Multiple-choice items are made up of a stem and answer options. The stem can be either a direct question or an incomplete sentence, and students must choose either the correct answer or the best answer.<sup>563</sup> Direct questions are considered better and less confusing stems, while best answer selection allows multiple-choice questions to be more challenging.

Multiple-choice questions offer a number of benefits to test creators. They are easy to score, particularly if electronic scanning of scores is used.<sup>564</sup> They can be used to assess both knowledge and affect.<sup>565</sup> They can be written at a wide variety of difficulty levels, and wrong answers can be analyzed to reveal students' misunderstandings.<sup>566</sup>

Multiple-choice questions also have a number of drawbacks. They cannot measure a student's ability to synthesize or be creative.<sup>567</sup> They are susceptible to guessing, since correct answers are provided. Guessing on multiple-choice

---

<sup>561</sup> Popham, Test Better, Teach Better. 85.

<sup>562</sup> Popham, Test Better, Teach Better. 79.

<sup>563</sup> Popham, Test Better, Teach Better. 80.

<sup>564</sup> Popham, Test Better, Teach Better. 81, 80.

<sup>565</sup> Popham, Test Better, Teach Better. 81.

<sup>566</sup> Popham, Test Better, Teach Better. 81.

<sup>567</sup> Popham, Test Better, Teach Better. 82.

tests can lead instructors to exaggerate students' learning.<sup>568</sup> Multiple-choice items are also less tolerant of poor test writing. For example, some test writers use answer choices that are obviously incorrect.<sup>569</sup> Also, negatively phrased items and "all of the above" answer choices can be unnecessarily confusing.<sup>570</sup> Unclear directions, unintentional clues, difficult vocabulary, and complex phrasing can all interfere with accurate assessments.<sup>571</sup> This last is true of all selected-response questions, and not just multiple-choice item types. Indeed, there is no research indicating that one type of selected-response item is better than the others.<sup>572</sup>

### **Standards-Based Tests**

In light of the many difficulties associated with traditional outcomes-based surveys and standardized tests, some educators call for "untraditional standardized achievement tests—ones that do not rely on norm-referenced comparisons but that still provide the evaluative evidence that the architects of educational accountability programs demand."<sup>573</sup> Instead, criterion-based tests

---

<sup>568</sup> Popham, Test Better, Teach Better. 81-82.

<sup>569</sup> Popham, Test Better, Teach Better. 82.

<sup>570</sup> Popham, Test Better, Teach Better. 83, 84.

<sup>571</sup> Popham, Test Better, Teach Better. 64.

<sup>572</sup> Popham, Test Better, Teach Better. 64.

<sup>573</sup> Popham, Test Better, Teach Better. 141.

are needed. Criterion-based tests reference a score back to a defined skill or body of knowledge, not to other students' performances.<sup>574</sup> Popham explains:

The kind of instructionally supportive standardized achievement test educators need to both boost instructional quality and supply accountability evidence must be constructed to provide meaningful criterion-referenced interpretations. This is essential, because it means that test developers are off the hook when it comes to score-spread. If no one is interested in comparing test takers with the norm group, then it's no longer necessary to include the score-spreading items.<sup>575</sup>

However, truly criterion-referenced tests are elusive. Tests that purport to be criterion-referenced are usually constructed in the same way as traditional standardized tests, with the same focus on score spread, and are therefore still "unsuitable for evaluating educational quality."<sup>576</sup> According to Crighton:

In practice...when constructing criteria for a criterion-referenced test, norm-referencing is unavoidable. Hidden behind each criterion is norm-referenced data: assumptions about how the average child in that particular age group can be expected to perform. Pure criterion-referenced assessment is rare and it would be better to think of assessment as being a hybrid of norm- and criterion-referencing. The same is true of setting standards, especially if they have to be reachable by students of varying ability: one has to know something about the norm before one can set a meaningful standard.<sup>577</sup>

Indeed, Popham explains that tests themselves are not norm-referenced or criterion-referenced. What is norm-referenced or criterion-referenced is the inferences educators make from the results.<sup>578</sup> Popham writes:

---

<sup>574</sup> Popham, Test Better, Teach Better. 141.

<sup>575</sup> Popham, Test Better, Teach Better. 141-142.

<sup>576</sup> Popham, Test Better, Teach Better. 140.

<sup>577</sup> Crighton, "Standardized Tests and Educational Policy." 2530.

<sup>578</sup> Popham, Test Better, Teach Better. 141.



A serious shortcoming of today's so-called standards-based tests and the whole standards-based reform strategy—is that these tests typically do not supply teachers with a report regarding a student's standard-by-standard mastery. How can teachers decide what aspects of their instruction need to be modified if they are unable to determine which content standards their students have mastered and which they have not? Without per-standard reporting, all that teachers get is a general and potentially misleading report of the student's overall standards mastery. This information has little instructional value.<sup>579</sup>

On the other hand, standard-by-standard reporting for individual students would be helpful to teachers, parents, and students.<sup>580</sup> Such reporting would require tests to include a sufficient number of questions per content standard to form accurate assessments of students' abilities, and would probably result in fewer standards being assessed.<sup>581</sup> However, "it is better for tests to measure a handful of powerful skills accurately than it is for tests to do an inaccurate job of measuring many skills."<sup>582</sup>

To be a meaningful assessment option, criterion-referenced tests need to be based on more explicit standards. Crighton points out, "Standards still tend to be expressed in terms of content covered and hours on the timetable...for each subject rather than student outcomes. When outcomes are mentioned, it is often in unmeasurable terms."<sup>583</sup> Popham explains this problem in more detail:

Another instructional shortcoming of most standards-based tests is that they don't spell out what they're actually measuring with sufficient clarity so that the teacher can teach toward the bodies of skills and knowledge

---

<sup>579</sup> Popham, Test Better, Teach Better. 143.

<sup>580</sup> Popham, Test Better, Teach Better. 143.

<sup>581</sup> Popham, Test Better, Teach Better. 143.

<sup>582</sup> Popham, Test Better, Teach Better. 143.

<sup>583</sup> Crighton, "Standardized Tests and Educational Policy." 2531-2532.

the tests represent.... For purposes of a teacher's instructional decision making, the difficulty is that the description of what the standardized achievement test measures is typically way too skimpy to help a teacher direct instruction properly.... Each of the high-priority content standards assessed must be accompanied by an assessment description that sets forth...just what a student needs to be able to do cognitively to perform well on the items measuring a particular content standard.<sup>584</sup>

He goes on to suggest that these assessment descriptions should be accompanied by illustrative examples.<sup>585</sup> Crighton remarks that fleshing out standards to this level would be useful, not only for outcomes-based survey and test approaches, but also for performance assessments.<sup>586</sup>

### ***Surveys and Tests of Information Literacy Skills***

Outcomes-based surveys abound in library instruction literature. Both Shonrock<sup>587</sup> and Merz and Mark<sup>588</sup> have published collections of information literacy surveys and tests, and Merz and Mark report that fifty percent of libraries surveyed use tests to assess instruction.<sup>589</sup>

Several California State University campuses have initiated survey and test measures of information literacy.<sup>590</sup> For example, California State University-Pomona developed a web-based multiple-choice test to "establish a baseline and

---

<sup>584</sup> Popham, Test Better, Teach Better. 143, 145.

<sup>585</sup> Popham, Test Better, Teach Better. 145.

<sup>586</sup> Crighton, "Standardized Tests and Educational Policy." 2531-2532.

<sup>587</sup> Shonrock, Evaluating Library Instruction.

<sup>588</sup> Merz and Mark, Clip Note #32.

<sup>589</sup> Merz and Mark, Clip Note #32. 8.

<sup>590</sup> Lorie Roth, "Educating the Cut-and-Paste Generation," Library Journal 124.18 (1999). 44.

benchmark measure of levels of entry-level information competence and computer literacy for all students at the university” and to “promote the development and implementation of an information competence policy and plan” for their campus.<sup>591</sup> California State University test creators advocated a web-based test in order to make test changes, data analysis, and distribution easier and to minimize costs.<sup>592</sup>

At the University of California-Berkeley campus, librarians developed a thirty-six question multiple-choice tool to test “lower-order” information literacy skills.<sup>593</sup> This test was used between 1994 and 1999 and had several goals. Generally, the test was intended to “establish a baseline of students skills around which an information literacy program might be built; to assess the effectiveness of particular library instruction sessions or approaches to instruction; to determine the impact of library instruction programs on student information literacy skills and academic success; and to generate data with which to communicate with faculty.”<sup>594</sup> The test was also part of a marketing plan. The librarians hoped to determine the skill levels of graduating seniors, assuming that if skill levels were low, the library could use the data to argue for a systematic program of library

---

<sup>591</sup> Kathleen Dunn, Information Competence Assessment: Web-Based Assessment of University Entry-Level Information Competence, 1999, April 30 2005  
<<http://www.csupomona.edu/~library/InfoComp/>>.

<sup>592</sup> Dunn, Information Competence Assessment.

<sup>593</sup> Patricia Davitt Maughan, "Assessing Information Literacy among Undergraduates: A Discussion of the Literature and the University of California-Berkeley Assessment Experience," College and Research Libraries 62.1 (2001).

<sup>594</sup> Maughan, "Assessing Information Literacy among Undergraduates."

instruction.<sup>595</sup> The test was administered to groups of graduating seniors three times, and the median score was a failing score.<sup>596</sup> Maughan states that the fundamental conclusion that can be made based on this survey is that “students think they know more about accessing information and conducting library research than they are able to demonstrate.”<sup>597</sup>

At the University of the Pacific, librarians used an outcomes-based test to measure effectiveness of a library tutorial. Librarians wanted “to incorporate interactive assessment tools into our Web-based tutorial and gather data on students’ attitudes about it.”<sup>598</sup> To do so, librarians developed a multiple-choice online exercise intended to measure student learning of the information contained in each tutorial section and placed it at the end of their web tutorial.<sup>599</sup> Overall, students scored well on this test. Ninety-five percent could decipher bibliographic records and locate periodical databases on the home page.<sup>600</sup> Only 54 percent could locate journal holdings in the catalog, and 24 percent could determine whether the full text of an article in a database was available.<sup>601</sup> Knight acknowledges that this test focuses only on “very basic skills.”<sup>602</sup>

---

<sup>595</sup> Maughan, "Assessing Information Literacy among Undergraduates."

<sup>596</sup> Maughan, "Assessing Information Literacy among Undergraduates."

<sup>597</sup> Maughan, "Assessing Information Literacy among Undergraduates."

<sup>598</sup> Knight, "The Role of Assessment in Library User Education."

<sup>599</sup> Knight, "The Role of Assessment in Library User Education."

<sup>600</sup> Knight, "The Role of Assessment in Library User Education."

<sup>601</sup> Knight, "The Role of Assessment in Library User Education."

<sup>602</sup> Knight, "The Role of Assessment in Library User Education."

Librarians in other countries have developed outcomes-based surveys and tests too. In Quebec, several universities collaborated to survey the information literacy skills of incoming students. These universities had two project goals. First, they wanted to determine students' skill levels in order to provide better services, and second, they wanted to provide libraries with better data so that they could argue for inclusion of information literacy skills in the university curriculum.<sup>603</sup> In a 2003 report, librarian and library school collaborators describe the 20-question multiple-choice survey that covers concept identification, search strategy, document types, search tools, and use of results.<sup>604</sup> Problem areas identified by the survey include difficulties removing non-significant words from search terms, incorrect use of Boolean "or", lack of knowledge of controlled vocabulary, inability to describe scholarly journals, failure to distinguish article databases from library catalogs, and difficulties with source citations.<sup>605</sup>

The Council of Australian University Librarians (CAUL) has developed two multiple-choice survey instruments.<sup>606</sup> Using these surveys, librarians found that "scholars use information differently in different discipline contexts."<sup>607</sup> As a result, CAUL plans to develop information literacy surveys for students in

---

<sup>603</sup> Diane Mittermeyer and Diane Quirion, Information Literacy: Study of Incoming First-Year Undergraduates in Quebec (National Library of Quebec, 2003). 5.

<sup>604</sup> Mittermeyer and Quirion, Information Literacy: Study of Incoming First-Year Undergraduates. 6.

<sup>605</sup> Mittermeyer and Quirion, Information Literacy: Study of Incoming First-Year Undergraduates.

<sup>606</sup> Council of Australian University Librarians, CAUL Information Skills Survey, 2004 <<http://www.anziil.org/news/february2004.htm>>.

<sup>607</sup> Council of Australian University Librarians, CAUL Information Skills Survey.

different disciplines, rather than attempting to use one survey to study all students.<sup>608</sup>

Slightly less common in library instruction literature is the outcomes-based standardized test. In 1992, Blandy reported that testing corporations believed “library skills cannot be covered by a standardized test.”<sup>609</sup> However, three notable library-originated examples exist. The Bay Area Community Colleges in California developed a 47-item multiple-choice and matching test, which is paired with a performance-based test described in Chapter 4.<sup>610</sup> The two-part test is intended to measure “information competence commensurate with a two-year college degree” and allow students to test out of a graduation requirement.<sup>611</sup> The test is not intended to reflect the needs of all California community colleges, only the San Francisco Bay Area schools.<sup>612</sup> Test developers focused on ACRL Standards 1, 2, 3, and 5 and strived to create a criterion-referenced test based on these standards.<sup>613</sup> The test, designed to take less than 3 hours, was field tested twice.<sup>614</sup> After each field test, items were revised. Test authors noted,

---

<sup>608</sup> Council of Australian University Librarians, CAUL Information Skills Survey.

<sup>609</sup> Susan Griswold Blandy, "The Librarians' Role in Academic Assessment and Accreditation: A Case Study," Assessment and Accountability in Reference Work (New York Haworth Press, 1992). 80.

<sup>610</sup> Topsy N. Smalley, Bay Area Community Colleges Information Competency Assessment Project, 2003, November 17 2003 <<http://www.topsy.org/ICAP/ICAPProject.html>>.

<sup>611</sup> Andy Kivel, "SF Bay Area Community College Information Competency Assessment Project," ACRL National Conference (Charlotte, North Carolina: Association of College and Research Libraries, 2003).

<sup>612</sup> Smalley, Bay Area Community Colleges.

<sup>613</sup> Smalley, Bay Area Community Colleges.

<sup>614</sup> Smalley, Bay Area Community Colleges.

“Occasionally there was disagreement about the correct answer of the test item.”<sup>615</sup> Even after revisions, some of the test questions do not follow “best practices” of test item creation. For example, some items include the use of negatives or have more than one possible right answer. Smalley admits that the validity and reliability of the test cannot yet be confirmed.<sup>616</sup>

UCLA developed a second outcomes-based standardized test worth noting. This test strives to obtain “an objective measure of [students’] information competence” and the use of library resources, online searching, and information-seeking concepts in order to improve library instruction.<sup>617</sup> Generally, students did not score well on the test; however, students who reported frequent library use, seniors, and humanities majors scored higher than other students.<sup>618</sup>

Librarians at UCLA, reflecting on their experiences with this test, acknowledge a few weaknesses. They note that it is difficult to balance coverage of the information literacy content the number of questions on a test.<sup>619</sup> For example, some of the outcomes the test seeks to assess were represented by multiple test items, some with one item, and some with none.<sup>620</sup> However, they note that adding more questions would have decreased the number of students

---

<sup>615</sup> Smalley, Bay Area Community Colleges.

<sup>616</sup> Smalley, Bay Area Community Colleges.

<sup>617</sup> Patti S. Caravello, Eloisa Gomez Borah, Judith Herschman and Eleanor Mitchell, Information Competence at UCLA: Report of a Survey Project, 2001, April 29 2005 <<http://www.library.ucla.edu/infocompetence/toc.html>>.

<sup>618</sup> Caravello, Borah, Herschman and Mitchell, Information Competence at UCLA.

<sup>619</sup> Caravello, Borah, Herschman and Mitchell, Information Competence at UCLA.

<sup>620</sup> Caravello, Borah, Herschman and Mitchell, Information Competence at UCLA.30.

who volunteered to take the test.<sup>621</sup> Furthermore, they point out that some outcomes simply do not lend themselves to assessment by a standardized instrument, such as outcomes that focus on the use or integration of information.<sup>622</sup> Second, UCLA librarians note that the score range on the test was 27 to 89 percent and scores fell in a normal curve.<sup>623</sup> However, they found that a few of the items on the test were not predictive of information competence.<sup>624</sup> Finally, they acknowledge the limitations of any multiple-choice test. They write:

The short multiple-choice test has its limits for assessing a large and complex set of knowledge and skill. Ideas for completely different types of testing should be considered, in which the student would actively demonstrate competence with research strategy concepts and tools rather than passively pick from given choices on a test. This might give a better, more comprehensive impression of information competence and where the weaknesses and misunderstandings lie. Such testing might take the form of live sessions...or open-ended questions where students have to come up with their own research strategies.<sup>625</sup>

The most well known outcomes-based test for library instruction is the Standardized Assessment of Information Literacy Skills (SAILS), which is still being developed. The authors of the test found that library instruction literature “does not contain or make reference to an instrument that is suitable for

---

<sup>621</sup> Caravello, Borah, Herschman and Mitchell, Information Competence at UCLA.30.

<sup>622</sup> Caravello, Borah, Herschman and Mitchell, Information Competence at UCLA.30.

<sup>623</sup> Caravello, Borah, Herschman and Mitchell, Information Competence at UCLA.30.

<sup>624</sup> Caravello, Borah, Herschman and Mitchell, Information Competence at UCLA.30.

<sup>625</sup> Caravello, Borah, Herschman and Mitchell, Information Competence at UCLA.30.



standardized, longitudinal, and cross-institutionally administered assessment.”<sup>626</sup>

The SAILS test is ambitiously conceptualized to fill this void as an “instrument for programmatic level assessment of information literacy skills that is valid and thus credible to university administrators and other academic personnel.”<sup>627</sup> Test creators “envision a standardized tool that is valid and reliable; contains items not specific to a particular institution or library but rather assesses at an institutional level; is easily administered; and provides for both external and internal benchmarking.”<sup>628</sup> It is hoped that the SAILS test will help “gather national data, provide norms, and compare information literacy measures with other indicators of student achievement” and that libraries using the test will be able to “pinpoint areas for improvement, identify and justify resource needs, assess and demonstrate effects of changes in their instructional programs.”<sup>629</sup> Test creators envision “combining pre- and post-testing with experimental and control conditions to answer the questions, ‘Does library instruction make a difference on campus?’ and ‘Does library instruction lead to the acquisition of information literacy skills?’”<sup>630</sup>

---

<sup>626</sup> Lisa G. O'Connor, Carolyn J. Radcliff and Julie A. Gedeon, "Applying Systems Design and Item Response Theory to the Problem of Measuring Information Literacy Skills," College and Research Libraries 63.6 (2002). 529.

<sup>627</sup> Peggy L. Maki, Assessing for Learning: Building a Sustainable Commitment across the Institution (Sterling, Virginia: Stylus, 2004). A-1.

<sup>628</sup> Maki, Assessing for Learning. A-1.

<sup>629</sup> Maki, Assessing for Learning. A-1.

<sup>630</sup> Lisa G. O'Connor, Carolyn J. Radcliff and Julie A. Gedeon, "Assessing Information Literacy Skills: Developing a Standardized Instrument for Institutional and Longitudinal Measurement," Association of College and Research Libraries 10th National Conference (Denver, Colorado: 2001). 163-164.

Development of this test commenced in 1998 at Kent State University when academic librarians began writing multiple-choice items to address ACRL's Objectives for Information Literacy Instruction: A Model Statement for Academic Librarians.<sup>631</sup> In 2002, Kent State University librarians received grant funding from the Institute of Museum and Library services to pilot the test at other institutions in three phases and ARL included it as a project in their New Measures Initiative.<sup>632</sup> In January 2003, ARL took over the coordination and management of the SAILS project.<sup>633</sup>

Results of the SAILS test are reported according to ACRL standards and by the nine skill sets established by SAILS developers.<sup>634</sup> During the development phases of the test, only average student scores are reported, rather than individual student scores.<sup>635</sup> Average student scores for an individual institution are also graphed against averages across participating institutions.<sup>636</sup> For example, North Carolina State University, a Phase II participant in SAILS development, learned that the average student who took the SAILS test at NCSU performed on all standards and all skill sets "at about the same level as the average student from all institutions combined."<sup>637</sup> This type of feedback has two

---

<sup>631</sup> Maki, Assessing for Learning. A-1.

<sup>632</sup> Maki, Assessing for Learning. A-1.

<sup>633</sup> Maki, Assessing for Learning. A-1.

<sup>634</sup> Maki, Assessing for Learning. 1.

<sup>635</sup> Maki, Assessing for Learning. 1.

<sup>636</sup> Maki, Assessing for Learning. 1.

<sup>637</sup> Maki, Assessing for Learning. 1.

limitations. First, while the test creators intend the test to be criterion-referenced, the feedback compares average students from multiple institutions and therefore appears norm-referenced. Second, because only general feedback is provided, it is difficult to use the SAILS test results for instructional decision-making.

Other limitations of the SAILS test include content coverage, item formats, and motivation issues. SAILS creators acknowledge that one of the greatest challenges of development is the “exceptional breadth and depth of the information literacy construct” and state that it is difficult to measure this construct using standardized assessment.<sup>638</sup> O’Connor notes that a standardized test that covers information literacy as it is conceptualized by the ACRL would require “hundreds of questions on a test that would require hours to complete.”<sup>639</sup> SAILS paper test forms include 30-35 questions and take an average of 45 minutes to complete. As a result, the SAILS test does not cover the full range of content represented by information literacy.

The questions included on the SAILS test are another area for examination. To write test questions, developers used Item Response Theory. Among other things, this theory helps developers produce score-spread to differentiate student test performance. However, the use of score-spread techniques results in norm-referencing, even when developers intend to create a criterion-referenced test. According to Popham, tests that purport to be criterion-

---

<sup>638</sup> Lisa O'Connor, "Project SAILS: Clearer Horizons for Information Literacy Assessment," LOEX Quarterly 31.1 (2004). 8.

<sup>639</sup> O'Connor, "Project SAILS." 8.

referenced but are constructed in the same way as traditional standardized tests are still “unsuitable for evaluating educational quality.”<sup>640</sup>

An additional limitation results from the web-based delivery format of the test. Because of the limitations of this format, test creators constrained item types to selected-response options, including multiple-choice and true/false.<sup>641</sup> To create selected-response test items, SAILS test creators suggest breaking learning outcomes into “manageable chunks”.<sup>642</sup> This is a difficult task. Information literacy is a complex concept that hinges on higher order thinking skill, and higher order thinking skills are difficult to assess using selected response questions. As a result, many SAILS test questions (see Figure 3.4)<sup>643</sup> are confusing or have multiple correct answers. SAILS test creators acknowledge the difficulty of writing good test questions.<sup>644</sup>

---

<sup>640</sup> Popham, Test Better, Teach Better. 140.

<sup>641</sup> Consuela Askew Waller, Lisa O'Connor, Carolyn Radcliff, Nan Seamans and Chris Sheetz, "Two (or Twelve) Heads Are Better Than One: Addressing the Complexities of Working with ACRL's Information Literacy Outcomes to Develop Assessment Tools," Association of Research Libraries 11th National Conference (Charlotte, North Carolina: 2003).

<sup>642</sup> Waller, O'Connor, Radcliff, Seamans and Sheetz, "Two (or Twelve) Heads Are Better Than One: Addressing the Complexities of Working with ACRL's Information Literacy Outcomes to Develop Assessment Tools."

<sup>643</sup> O'Connor, Radcliff and Gedeon, "Applying Systems Design and Item Response Theory." 534.

<sup>644</sup> O'Connor, Radcliff and Gedeon, "Assessing Information Literacy Skills: Developing a Standardized Instrument for Institutional and Longitudinal Measurement." 170-171.

**REVISION 3 (After field trials)**

Academic libraries are generally thought of as collections of materials in print and electronic formats. Some of these materials are made available to users through the Web but are not included in what we traditionally think of as the Web.

The World Wide Web is a means of communication. Computers all over the world network with one another by using a common language.

Which of the following statements are generally true about academic libraries and/or the Web?

Put a **W** if the statement is true about the Web.

Put an **L** if the statement is true about the academic library.

Put a **B** if the statement is true about both the academic library and the Web.

Put a **N** if the statement is **not** true about either the academic library or the Web.

\_\_\_ All its resources are free and accessible to students.

\_\_\_ Anyone can add information to it.

\_\_\_ Targets all audiences, including shoppers, support groups, scholars, students, hobbyists, businesses.

\_\_\_ Has materials that have been purchased on behalf of students.

\_\_\_ Information must have been deemed authoritative to be included.

\_\_\_ Is organized systematically with a classification scheme.

\_\_\_ Offers online option to ask questions.

*Figure 3.4. Sample SAILS Test Question*

A final reservation regarding the assessment of information literacy skills using a standardized test centers on student motivation. It is unclear how student motivation may be impacted by standardized assessments of information literacy. In the development phases, students who take the SAILS test "are either being required to complete the instrument outside of class or enticed to do so through extra credit or other type of incentive."<sup>645</sup> Extrinsic rewards and requirements have convinced students to take the SAILS test in its development stages, but it is difficult to determine how much effort students will exhibit on a finalized test of this type. More research is needed to explore students'

<sup>645</sup> O'Connor, "Project SAILS." 8.

motivation levels on standardized approaches to information literacy assessment, particularly when no high-stakes decisions hinge on the results.

### **Outcomes-Based Performance Assessments**

Many modern educators feel a “growing dissatisfaction with selected-response testing.”<sup>646</sup> Wiggins writes, “Circling ‘correct’ answers to problems only test makers care about is not ‘knowing,’ nor it is the aim of teaching.”<sup>647</sup>

Battersby suggests that “standard written tests [are] seldom...appropriate to assess outcomes—because what students will do with the knowledge and skills they are learning is not usually writing tests or essays.”<sup>648</sup>

Because of the limitations of outcomes-based surveys and standardized tests, the emphasis on quantitative results that once dominated assessment conversations in higher education is beginning to give way to discussions of qualitative forms of assessment.<sup>649</sup> Marzano, Pickering, and McTighe explain the impetus for recent changes:

A revolution in assessment is necessary given (1) the changing nature of educational goals to encompass a broad array of academic and non academic competencies, (2) the need for assessment practices to enhance the learning and teaching processes, and (3) the need for record-keeping and reporting systems to provide accurate and useful information concerning students’ mastery of specific knowledge and skills. Indeed,

---

<sup>646</sup> Edward A. Silver, "Performance Assessment," Encyclopedia of Education, ed. James W. Guthrie, vol. 1 (New York: Macmillan Reference USA, 2003). 135.

<sup>647</sup> Wiggins, "Creating Tests Worth Taking." V-6: 8.

<sup>648</sup> Battersby, "So, What's a Learning Outcome Anyway?."

<sup>649</sup> Pausch and Popp, "Assessment of Information Literacy."

such a revolution is currently underway in the form of an emphasis on performance assessment.<sup>650</sup>

This “revolutionary” approach to assessment focuses on creating assignments and assessments that ask students to perform in ways that simulate real life uses of knowledge and skills.<sup>651</sup> These performance approaches to assessment reinforce the concept that what students learn in class should be usable in life outside of class.<sup>652</sup> Farmer explains, “It’s the difference between describing how to ride a bike and actually putting the foot to the pedal and pumping down the street. Thus a scantron ‘bubble’ test would be an unlikely...assessment tool.”<sup>653</sup>

According to Shavelson and Baxter, outcomes-based performance assessments “rest on a set of assumptions about effective and responsible teaching that differ significantly from assumptions of the past.”<sup>654</sup> In the performance assessment paradigm, learning is conceptualized as an active process where “students construct meaning and knowledge: they do not have meaning or knowledge handed to them in a book or lecture. Learning, then, is a process of students ‘making sense’ of how things fit together; factual and procedural knowledge is built along the way.”<sup>655</sup> In performance assessments, students “are given opportunities to demonstrate their understanding and to

---

<sup>650</sup> Marzano, Pickering and McTighe, Assessing Student Outcomes. 12.

<sup>651</sup> Battersby, "So, What's a Learning Outcome Anyway?."

<sup>652</sup> Battersby, "So, What's a Learning Outcome Anyway?."

<sup>653</sup> Farmer, "Authentic Assessment of Information Literacy."

<sup>654</sup> Richard J. Shavelson and Gail P. Baxter, "Linking Assessment with Instruction," A Handbook for Student Performance in an Era of Restructuring, eds. Robert E. Blum and Judith A. Arter (Alexandria, Virginia: Association for Supervision and Curriculum Development, 1996). IV-7: 1.

<sup>655</sup> Shavelson and Baxter, "Linking Assessment with Instruction." IV-7: 1.

thoughtfully apply knowledge, skills, and habits of mind in a variety of contexts” as they would in the real world.<sup>656</sup>

There are a number of ways to structure performance assessment. The important thing is to create “assignments and assessments that simulate as much as possible the situations in which students would make integrated use of the knowledge, skills, and values developed in the course.”<sup>657</sup> Real world scenarios demonstrate that students should be able to use their learning outside the classroom.<sup>658</sup> For example, instead of scoring a student’s answers to a set of answers provided by a teacher,<sup>659</sup> a performance assessor might observe a student’s performance of a task or a product of a performance, and judge its quality.<sup>660</sup> In a performance assessment situation, there may be one or more assessors, the task may be carried out under controlled or real life conditions,<sup>661</sup> and judgments of quality of work are based on agreed-upon criteria.<sup>662</sup> Other methods that have been used for performance assessments are open-ended or extended response exercises (such as questions or other prompts that ask

---

<sup>656</sup> Marzano, Pickering and McTighe, Assessing Student Outcomes. 13.

<sup>657</sup> David Sweet, "Performance Assessment," Education Research Consumer Guide 2 (1993).

<sup>658</sup> Mark Battersby, So, What's a Learning Outcome Anyway? Learning Outcomes and the Learning Paradigm, 20022004  
<<http://merlin.capcollege.bc.ca/mbatters/whatsalearningoutcome.htm>>.

<sup>659</sup> Sweet, "Performance Assessment."

<sup>660</sup> Silver, "Performance Assessment." 134.

<sup>661</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment. 4.

<sup>662</sup> Sweet, "Performance Assessment."



students to explore a topic in writing), extended tasks, or portfolios.<sup>663</sup>

Constructed-response tasks have become popular because they capture what is valued instructionally in a form that can be easily assessed.<sup>664</sup> In general, “Most educators regard performance assessment as an attempt to measure a student’s mastery of a high-level, sometimes quite sophisticated skill through the use of fairly elaborate constructed-response items and a rubric.”<sup>665</sup> Many students consider constructed-response tasks that test both understanding and the ability to transfer learning quite challenging.<sup>666</sup> Battersby suggests that students’ reactions “underline the difficulty of true understanding...and the need to teach for transfer and application.”<sup>667</sup>

Performance assessments should meet a number of goals. Performance assessments should be meaningful and authentic.<sup>668</sup> Wiggins states that performance assessments should involve actual “performances, not drills. A test of many items (a drill) is not a test of knowledge in use. ‘Performance’ is not just doing simplistic tasks that cue us for the desired bit of knowledge. It entails ‘putting it all together’ with good judgment; good judgment cannot be tested

---

<sup>663</sup> Sweet, "Performance Assessment."

<sup>664</sup> Silver, "Performance Assessment." 135.

<sup>665</sup> Popham, Test Better, Teach Better. 101.

<sup>666</sup> Battersby, So, What's a Learning Outcome Anyway? Learning Outcomes and the Learning Paradigm.

<sup>667</sup> Battersby, So, What's a Learning Outcome Anyway? Learning Outcomes and the Learning Paradigm.

<sup>668</sup> Wiggins, "Creating Tests Worth Taking." V-6: 3.

through isolated, pat drills.”<sup>669</sup> Shepard agrees that assessments should “require more complex and challenging mental processes from students. They should acknowledge more than one approach or one right answer and should place more emphasis on un-coached explanations and real student products.”<sup>670</sup> Shepard also notes that performance assessments should be open-ended enough to allow each student “to bring to it his individual gifts and to maximize individual learning.”<sup>671</sup> Shavelson and Baxter expand on the need for assessments with more flexibility and more than one right answer. They explain:

Tests should contain tasks for which there are alternative solutions. Hence, the tasks confronting students should be holistic in nature; the amount of time to solve them will exceed the usual thirty seconds allocated to an exercise provided by the teacher or a multiple-choice question on a test. Evaluation of performance on such tasks should capture the diversity of problem-solving strategies, compare them on a common metric, and provide credit for partial knowledge and well-reasoned, even if somewhat erroneous, solution strategies.<sup>672</sup>

Sweet states that the ability for students to “actively develop their approaches to the task under defined conditions, knowing that their work will be evaluated according to agreed upon standards” is what distinguishes performance assessments from traditional testing.<sup>673</sup>

Among other uses, performance assessments can provide a means for “assessing student learning outcomes...assigning course grades, communicating

---

<sup>669</sup> Wiggins, "Creating Tests Worth Taking." V-6: 4.

<sup>670</sup> Shepard, "Why We Need Better Assessments." I-2: 4.

<sup>671</sup> Shepard, "Why We Need Better Assessments."

<sup>672</sup> Shavelson and Baxter, "Linking Assessment with Instruction." IV-7: 2.

<sup>673</sup> Sweet, "Performance Assessment."

expectations, providing feedback to students, and guiding instructional decisions.”<sup>674</sup> Generalizable performance assessments are the most useful.

Baker et. al. describe the importance of the generalizability of performance assessments:

Generalizability and dimensionality are critical to the use and interpretation of performance assessment, in part, because each individual assessment is time-consuming to administer and score. Consequently, it is important to gain as much valid information from the performances on each task as is possible. If an assessment has a high level of generalizability, one can use fewer tasks to assess the domain.... Evaluations of the level of generalizability across both tasks and raters are therefore critical to the overall evaluation of any performance assessment. The dimensionality question addresses the validity of the scoring rubric and the extent to which patterns of student performance follow predictions.<sup>675</sup>

### ***Constructed-Response Items***

Constructed-response items are test items that require students to respond by “constructing, that is, generating an answer, an essay, or whatever the item calls for.”<sup>676</sup> Constructed-response items offer many benefits. Popham states that the chief virtue of constructed-response items is that they “require students to create their own responses rather than select a pre-packaged response from the answer shelf. Clearly, creating a response represents a more complicated and difficult task.”<sup>677</sup> In fact, Popham goes so far as to state, “If you

---

<sup>674</sup> Silver, "Performance Assessment." 6.

<sup>675</sup> Eva L. Baker, Jamal Abedi, Robert L. Linn and David Niemi, "Dimensionality and Generalizability of Domain-Independent Performance Assessments," Journal of Education Research 89.4 (1995). 198.

<sup>676</sup> Popham, Test Better, Teach Better. 63.

<sup>677</sup> Popham, Test Better, Teach Better. 86.

want your students to master truly powerful cognitive skills, you almost always need to rely on at least a certain number of constructed-response items.”<sup>678</sup> McBride offers a number of additional benefits of constructed-response items. She notes that constructed-response items elicit student responses that more closely mirror real life skills, are relatively easy to create, and can be administered quickly.<sup>679</sup> She also states that constructed-response items test skills that cannot be assessed in other ways, such as organization and communication skills.<sup>680</sup> Finally, she suggests that constructed-response items are not as susceptible to guessing as selected-response test items.<sup>681</sup> Popham agrees, remarking that “Students who might stumble onto a selected-response item’s correct answer simply by casting a covetous eye over the available options would never be able to concoct an original correct answer without access to such options.”<sup>682</sup> In addition to eliminating student guessing, constructed-response items may contribute to valid assessments better than selected-response items.<sup>683</sup> This is because they require students to understand a concept well enough to create a response.<sup>684</sup> All of these benefits lead to the conclusion that

---

<sup>678</sup> Popham, Test Better, Teach Better. 88.

<sup>679</sup> Amanda McBride, "Letting Students Shine: Assessment to Promote Student Learning," Focus 31 (1999). 12-13.

<sup>680</sup> McBride, "Letting Students Shine." 12-13.

<sup>681</sup> McBride, "Letting Students Shine." 12-13.

<sup>682</sup> Popham, Test Better, Teach Better. 87.

<sup>683</sup> Popham, Test Better, Teach Better. 87.

<sup>684</sup> Popham, Test Better, Teach Better. 87.

constructed-response test items will allow both teachers and students to “reap giant instructional payoffs.”<sup>685</sup>

Constructed-response test items have several limitations. For example, constructed-response items can be time-consuming to score.<sup>686</sup> Usually, constructed-responses require human involvement in grading rather than machine grading.<sup>687</sup> Because constructed-response items gather higher-level cognitive answers from students, they take more time and greater attention to objective scoring techniques.<sup>688</sup> Attention must be given to strategies for avoiding evaluator bias so that tests are fair and valid.<sup>689</sup> In fact, Popham states that the “more complex the task that’s presented to students in a constructed-response item, the tougher it is for teachers to score.”<sup>690</sup> Two additional limitations of constructed-response items are that they may penalize students who do not read or write well<sup>691</sup> and they are “tougher” for test-takers.<sup>692</sup>

Increasingly, educators are adding constructed-response items into their assessment programs.<sup>693</sup> The most common constructed-response test items

---

<sup>685</sup> Popham, Test Better, Teach Better. 88.

<sup>686</sup> McBride, "Letting Students Shine." 12-13.

<sup>687</sup> Popham, Test Better, Teach Better. 87.

<sup>688</sup> Popham, Test Better, Teach Better. 64.

<sup>689</sup> McBride, "Letting Students Shine." 12-13.

<sup>690</sup> Popham, Test Better, Teach Better. 87.

<sup>691</sup> McBride, "Letting Students Shine." 12-13.

<sup>692</sup> Popham, Test Better, Teach Better. 87.

<sup>693</sup> Robert L. Johnson, James Penny and Belita Gordon, "The Relation between Score Resolution Methods and Interrater Reliability: An Empirical Study of an Analytic Scoring Rubric," Applied Measurement in Education 13.2 (2000). 122.

are short-answer items and essay items.<sup>694</sup> Short-answer items require students to supply “a word, a phrase, or a sentence or two in response either to a direct question or an incomplete statement.”<sup>695</sup> Like other constructed-response items, students must create an answer, not just select an answer from a series of choices. Short-answer items have the benefit of being time-efficient. They can be answered quickly and scored quickly.<sup>696</sup> Of course, lengthier short-answer items take more time to complete and score, but they also offer more content to score and thus more revealing responses.<sup>697</sup> Short-answer items offer another benefit: they provide “insight into [students’] understanding, revealing if they are on the mark or conceptualizing something very differently from how the teacher intended it to be understood.”<sup>698</sup>

By following a few rules, educators can achieve better results from short-answer test items. First, test writers should structure short-answer items so that they elicit short and unique responses.<sup>699</sup> When possible, they should be formatted as questions rather than as incomplete sentences.<sup>700</sup> Students should

---

<sup>694</sup> Popham, Test Better, Teach Better. 63.

<sup>695</sup> Popham, Test Better, Teach Better. 88.

<sup>696</sup> Popham, Test Better, Teach Better. 88.

<sup>697</sup> Popham, Test Better, Teach Better. 89.

<sup>698</sup> Popham, Test Better, Teach Better. 88.

<sup>699</sup> Popham, Test Better, Teach Better. 90.

<sup>700</sup> Popham, Test Better, Teach Better. 90.

be provided with sufficient space to write their answers and all answer spaces should be of the same length.<sup>701</sup>

Longer constructed-response items are often referred to as essay items. According to Popham, "As an assessment tactic to measure truly sophisticated types of student learning, essay items do a terrific job."<sup>702</sup> One of the greatest benefits of essay items is the flexibility they offer. Essays can be used to measure student achievement of many types of learning outcomes.<sup>703</sup> McBride points out several additional benefits of using essay items to assess student learning. Essays require students to use communication and reasoning skills, they assess complex and higher-level thinking skills, and they can reflect real world tasks.<sup>704</sup> Furthermore, they encourage students to see beyond one right answer, can be self- or peer-assessed, and may reveal in-depth student understanding.<sup>705</sup>

Of course, essay items have limitations as well. For example, they are time consuming to score.<sup>706</sup> Recent research in electronic scoring of student essays shows that this limitation may soon be counteracted by technological solutions.<sup>707</sup> Scoring of essay items can be subjective and susceptible to

---

<sup>701</sup> Popham, Test Better, Teach Better. 92, 91.

<sup>702</sup> Popham, Test Better, Teach Better. 92.

<sup>703</sup> Popham, Test Better, Teach Better. 92.

<sup>704</sup> McBride, "Letting Students Shine." 12-13.

<sup>705</sup> McBride, "Letting Students Shine." 12-13.

<sup>706</sup> McBride, "Letting Students Shine." 12-13.

<sup>707</sup> Popham, Test Better, Teach Better. 92-93.

evaluator bias, which impacts the validity and fairness of the test.<sup>708</sup> However, model answers and rubrics can prevent such subjectivity.<sup>709</sup> Indeed, “we have learned from efforts to evaluate essays in statewide and national tests that it is possible to do so with a remarkable degree of accuracy, provided that sufficient resources are committed to the effort and well trained scoring personnel are in place.”<sup>710</sup> Finally, essay items do not favor students with poor communication skills.<sup>711</sup>

Four rules govern essay item development. First, test creators should ask more questions requiring shorter answers instead of few questions asking for long answers; in this way the test can cover more content.<sup>712</sup> Essay test items should be written clearly and explicitly so that students have no difficulty understanding the prompt.<sup>713</sup> Students should also be provided with the point total for each question, an acceptable response length, and information about the time allotted to answer the item.<sup>714</sup> Finally, test creators should “gauge a question’s quality by creating a trial response.”<sup>715</sup>

---

<sup>708</sup> McBride, "Letting Students Shine." 12-13.

<sup>709</sup> McBride, "Letting Students Shine." 12-13.

<sup>710</sup> Popham, Test Better, Teach Better. 93.

<sup>711</sup> McBride, "Letting Students Shine." 12-13.

<sup>712</sup> Popham, Test Better, Teach Better. 95.

<sup>713</sup> Popham, Test Better, Teach Better. 93.

<sup>714</sup> Popham, Test Better, Teach Better. 94.

<sup>715</sup> Popham, Test Better, Teach Better. 95.



### ***Theoretical Background: Education***

Outcomes-based performance assessment is grounded in constructivist and social-constructivist educational theory. These student-centered pedagogical theories have as their central emphasis the learning processes of students, not the subject matter that instructors impart.<sup>716</sup> According to Elmborg, “the essential defining trait of [constructivist] theorists is an insistence that knowledge is ‘constructed’ by individuals rather than passed on fully-formed from teachers to students.”<sup>717</sup> Thus constructivist theories of education are in opposition to earlier objectivist models that are firmly entrenched in educational traditions.<sup>718</sup> Objectivism views knowledge as a “passive reflection of the external, objective reality. This implies a process of ‘instruction,’” rather than a process of learning.<sup>719</sup> In contrast to objectivism, Lamon describes constructivism as:

An epistemology, or a theory, used to explain how people know what they know. The basic idea is that problem solving is at the heart of learning, thinking, and development. As people solve problems and discover the consequences of their actions—through reflecting on past and immediate experiences—they construct their own understanding. Learning is thus an active process that causes a change in the learner. This is achieved through the activities the learner engages in, including the consequences of those activities, and through reflection. People only deeply understand what they have constructed.<sup>720</sup>

---

<sup>716</sup> Elmborg, "Teaching at the Desk." 457.

<sup>717</sup> Elmborg, "Teaching at the Desk." 457.

<sup>718</sup> Mary Lamon, "Constructivist Approach," *Encyclopedia of Education*, ed. James W. Guthrie, vol. 4 (New York: Macmillan Reference USA, 2003). 1463.

<sup>719</sup> Lamon, "Constructivist Approach." 1463.

<sup>720</sup> Lamon, "Constructivist Approach." 1463.

Shepard describes the advent of constructivist theory as a “cognitive revolution [that] reintroduced the concept of mind.”<sup>721</sup> As a result of this revolution, educators questioned the “mechanistic” theories of instruction and embraced the concept that learning is an active process in which students construct and make sense of ideas.<sup>722</sup> She explains:

From cognitive theory we have also learned that existing knowledge structures and beliefs work to enable or impede new learning, that intelligent thought involves self-monitoring and awareness about when and how to use new skills, and that ‘expertise’ develops in a field of study as a principled and coherent way of thinking and representing problems, not just as an accumulation of information.<sup>723</sup>

In constructivist environments, learning is promoted through active engagement and purposeful interaction in real world, authentic problem solving, critical thinking,<sup>724</sup> and knowledge creation.<sup>725</sup> Learning in this way is “integrated and complex”, rather than “sequential and linear.”<sup>726</sup> Gabler and Schroeder describe constructivist instructors as those who help learners connect new knowledge to old knowledge, act as a facilitator, guide students through a process of “cognitive restructuring” rather than memorizing facts, involve students in teaching and learning processes, and continually reflect on their

---

<sup>721</sup> Shepard, "The Role of Assessment in a Learning Culture." 6.

<sup>722</sup> Shepard, "The Role of Assessment in a Learning Culture." 6-7.

<sup>723</sup> Shepard, "The Role of Assessment in a Learning Culture." 6-7.

<sup>724</sup> Ina Claire Gabler and Michael Schroeder, Seven Constructivist Methods for the Secondary Classroom: A Planning Guide for Invisible Teaching (Boston: Allyn and Bacon, 2003). 4.

<sup>725</sup> Chandra J. Foote, Paul J. Vermette and Catherine F. Battaglia, Constructivist Strategies: Meeting Standards and Engaging Adolescent Minds (Larchmont, New York: Eye on Education, 2001). 24.

<sup>726</sup> Foote, Vermette and Battaglia, Constructivist Strategies. 24.

practices.<sup>727</sup> Lamon adds that the role of a constructivist teacher is to act as an “expert learner” rather than a lecturer.<sup>728</sup> Such teachers “guide students into adopting cognitive strategies such as self testing, articulating understanding, asking probing questions, and reflection.”<sup>729</sup> He adds, “The role of the teacher in constructivist classrooms is to organize information around big ideas that engage the students’ interest, to assist students in developing new insights, and to connect them with their previous learning.”<sup>730</sup> Norman and Spohrer explain that constructivist learning theory is learner-focused, problem-based, and intrinsically motivated:

At the heart is the idea that people learn best when engrossed in the topic, motivated to seek out new knowledge and skills because they need them in order to solve the problem at hand. The goal is active exploration, construction and learning rather than the passivity of lecture attendance and textbook reading. The major theme is one of focusing education around a set of realistic, intrinsically motivating problems. Students work to solve these problems.... Teachers carefully structure the problems so that in the course of solution, students naturally pass through and acquire all topics of relevance.<sup>731</sup>

Constructivist teaching approaches are also strongly connected to the use of technology in the classroom. Lamon reports that teachers who use information and communication technologies are more likely to take a

---

<sup>727</sup> Gabler and Schroeder, Seven Constructivist Methods. 4-5.

<sup>728</sup> Lamon, "Constructivist Approach." 1465.

<sup>729</sup> Lamon, "Constructivist Approach." 1465.

<sup>730</sup> Lamon, "Constructivist Approach." 1465.

<sup>731</sup> Donald A. Norman and James C. Spohrer, "Learner-Centered Education," Communications of the ACM 39.4 (1996). 26.

constructivist approach toward teaching and learning.<sup>732</sup> Furthermore, information and communication technology can “capture the cognitive processes learners engage in when solving problems.”<sup>733</sup> As a result, teachers have increased feedback that they can use as they reflect on their teaching practices and plan ways to aid their students’ deep learning.<sup>734</sup> Finally, pairing technology and constructivist approaches allows teachers to learn from each other.<sup>735</sup> Generally, the constructivist approach to teaching and learning is widely accepted by most researchers.<sup>736</sup>

The origins of constructivism are rooted in the writings of John Dewey, Jean Piaget, and Lev Vygotsky.<sup>737</sup> John Dewey contributed the idea that “knowledge is a collective enterprise and not the sole possession of the teacher...[and] that knowledge is constructed by individuals and not dispensed as a commodity.”<sup>738</sup> These concepts “significantly altered the relationship between teachers and their students and students and their learning.”<sup>739</sup> Dewey also believed that learning required students to connect new ideas with personal

---

<sup>732</sup> Lamon, "Constructivist Approach." 1465.

<sup>733</sup> Lamon, "Constructivist Approach." 1465.

<sup>734</sup> Lamon, "Constructivist Approach." 1465.

<sup>735</sup> Lamon, "Constructivist Approach." 1465.

<sup>736</sup> Lamon, "Constructivist Approach." 1466.

<sup>737</sup> Gabler and Schroeder, Seven Constructivist Methods. 4.

<sup>738</sup> Foote, Vermette and Battaglia, Constructivist Strategies. 13.

<sup>739</sup> Foote, Vermette and Battaglia, Constructivist Strategies. 13.

experience.<sup>740</sup> This concept is also a precept of constructivism. Constructivists believe, as Dewey did, that “meaning cannot be handed to a person; it must be provoked through activity and revolve around problems that are important—not to the teacher—but to the student.”<sup>741</sup> In fact, Dewey can be credited with the idea that schools should include real world problems and problem solving in teaching and learning.<sup>742</sup> Dewey also contributed the idea of action to constructivism. Dewey believed that active learning and problem solving should be central to learning.<sup>743</sup> He argued that “schooling was not preparation for life, it was life itself and, therefore, education shared an organic and inseparable connection with personal experience.”<sup>744</sup>

Jean Piaget also contributed important concepts to constructivist theory. Piaget believed that the learning process is transformative rather than cumulative.<sup>745</sup> As a result, he theorized that students do not gain small pieces of knowledge and compile them at some later time.<sup>746</sup> He argued instead that students try to make sense of what they know and what they learn from the start.<sup>747</sup> Lamon explains, “This understanding is progressively reformed as new

---

<sup>740</sup> Foote, Vermette and Battaglia, Constructivist Strategies. 15.

<sup>741</sup> Foote, Vermette and Battaglia, Constructivist Strategies. 15.

<sup>742</sup> Lamon, "Constructivist Approach." 1464.

<sup>743</sup> Lamon, "Constructivist Approach." 1464.

<sup>744</sup> Foote, Vermette and Battaglia, Constructivist Strategies. 15.

<sup>745</sup> Lamon, "Constructivist Approach." 1463.

<sup>746</sup> Lamon, "Constructivist Approach." 1463.

<sup>747</sup> Lamon, "Constructivist Approach." 1463.

knowledge is acquired, especially new knowledge that is incompatible with their previous understanding.”<sup>748</sup>

Lev Vygotsky is credited with the theories that make up social-constructivism. Vygotsky believed that what students learn is socially and culturally determined and developed through “socially supported interactions”<sup>749</sup> such as “external and social activities, including communication, with more competent others.”<sup>750</sup> Thus social constructivism views learners as novices in a community<sup>751</sup> who must be acclimated and acculturated before becoming expert members of the community.

Interestingly, social constructivism is the dominant learning theory of academic writing programs.<sup>752</sup> In this theory, writing teachers act as mentors and guides who help students become members of the discourse of academia.<sup>753</sup> According to Elmborg, learning to conduct research is very similar to learning to write in an academic discipline.<sup>754</sup> He writes, “Both are based on publicly known and agreed upon assumptions about how college students should do their intellectual work.”<sup>755</sup> Elmborg suggests that librarians should help students

---

<sup>748</sup> Lamon, "Constructivist Approach." 1463.

<sup>749</sup> Shepard, "The Role of Assessment in a Learning Culture." 7.

<sup>750</sup> Lamon, "Constructivist Approach." 1463.

<sup>751</sup> Elmborg, "Teaching at the Desk." 457-458.

<sup>752</sup> Elmborg, "Teaching at the Desk." 457-458.

<sup>753</sup> Elmborg, "Teaching at the Desk." 458.

<sup>754</sup> Elmborg, "Teaching at the Desk." 458.

<sup>755</sup> Elmborg, "Teaching at the Desk." 458.

understand these assumptions and adopt a social constructivist model of instruction.<sup>756</sup>

Accepting a social constructivist approach to instruction has implications for the assessment of that instruction. Shepard argues that it makes sense to develop assessment practices that are consistent with social constructivist pedagogy. She writes:

If instructional goals include developing students' metacognitive abilities, fostering important dispositions, and socializing students into the discourse and practices of academic disciplines, then it is essential that classroom routines and corresponding assessments reflect these goals as well. This means expanding the armamentarium for data gathering to include observations, clinical interviews, reflective journals, projects, demonstrations, collections of student work, and students' self-evaluations, and it means that teachers must engage in systematic analysis of the available evidence.<sup>757</sup>

She suggests that school learning should be "authentic and connected to the world outside of school" to make learning and assessment interesting and motivating. Furthermore, Shepard states that "good assessment tasks are interchangeable with good instructional tasks," reinforcing the concept that instruction and assessment are inextricably linked (see Figure 3.5).

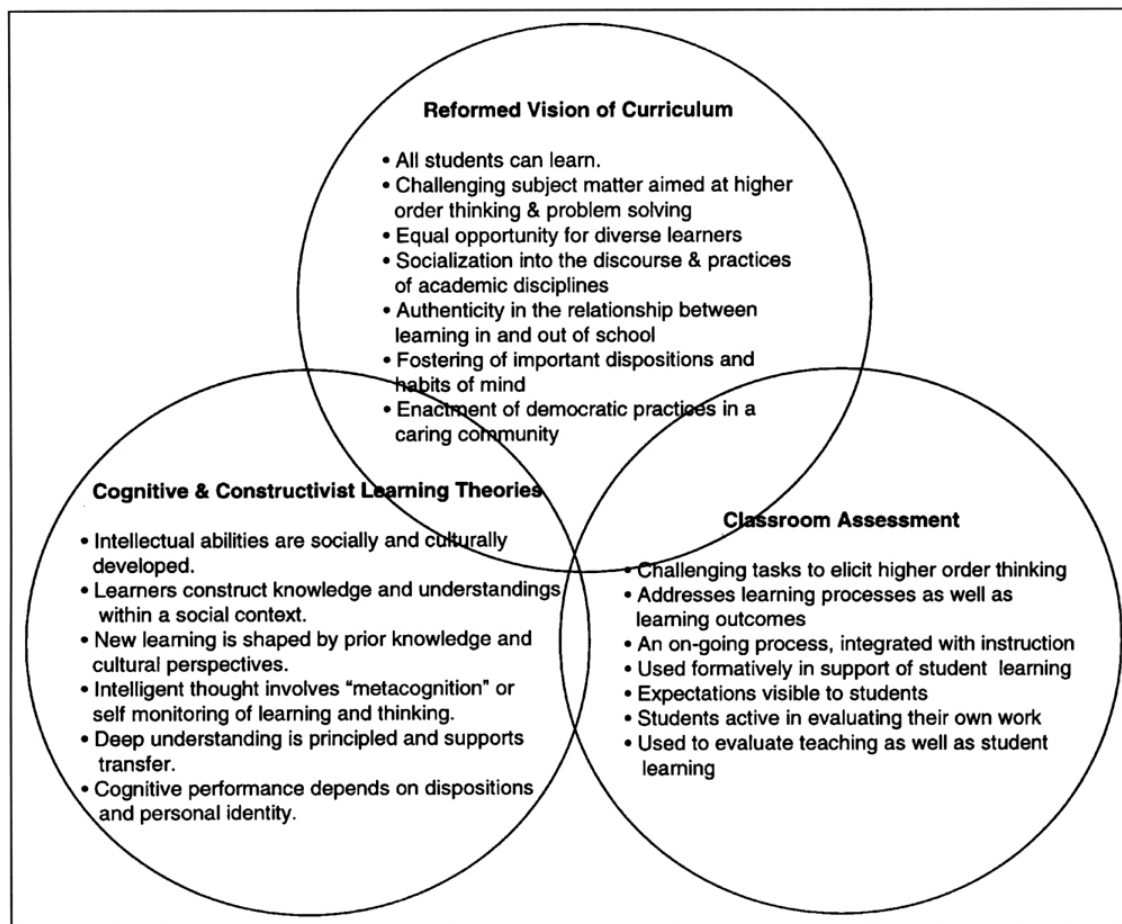
Shepard also argues that to support this social constructivist model of teaching, classroom assessment must change to "better represent important thinking and problem solving skills in each of the disciplines."<sup>758</sup> Shepard points out that assessment from a constructivist viewpoint does not match the goals of

---

<sup>756</sup> Elmborg, "Teaching at the Desk." 458.

<sup>757</sup> Shepard, "The Role of Assessment in a Learning Culture." 8.

<sup>758</sup> Shepard, "The Role of Assessment in a Learning Culture." 7.



*Figure 3.5. Educational Theory and Assessment*

many current tools used to evaluate student learning. Shepard writes:

We have not only to make assessment more informative, more insightfully tied to learning steps, but at the same time we must change the social meaning of evaluation. Our aim should be to change our cultural practices so that students and teachers look to assessment as a source of insight and help instead of an occasion for meting out rewards and punishments.<sup>759</sup>

Lamon agrees that the formative assessments that provide insight and help are more valuable to learners than summative assessments, but she too acknowledges that formative assessments do not always mesh well with

<sup>759</sup> Shepard, "The Role of Assessment in a Learning Culture." 10.



standards-based approaches to education.<sup>760</sup> Shepard further acknowledges the disconnect between social constructivist assessments and the traditional testing paradigm. She explains:

[The] emergent, constructivist paradigm in which teachers' close assessment of students' understandings, feedback from peers, and student self-assessments, would be a central part of the social processes that mediate the development of intellectual abilities, construction of knowledge, and formation of students' identities. The best way to understand dissonant current practices...is to realize that instruction (at least in its ideal form) is drawn from the emergent paradigm, while testing is held over from the past.<sup>761</sup>

### ***Theoretical Background: Motivation***

Not only does outcomes-based performance assessment suit the requirements of constructivist pedagogy, it is also well aligned with educational motivation theories. The study of motivation focuses on "what pushes or pulls an individual to start, direct, sustain, and finally end an activity."<sup>762</sup> Norman and Spohrer define motivated students as "engaged" students and state that "motivation...can make more of a difference between (National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development) success and failure than any other factor."<sup>763</sup>

Educators and assessors who concern themselves with student motivation face an uphill battle. Battersby believes that many students in higher education

---

<sup>760</sup> Lamon, "Constructivist Approach." 1465.

<sup>761</sup> Shepard, "The Role of Assessment in a Learning Culture." 4.

<sup>762</sup> Sandra Graham, "Motivation: Overview," Encyclopedia of Education, ed. James W. Guthrie, vol. 5 (New York: Macmillan Reference USA, 2003). 1690.

<sup>763</sup> Norman and Spohrer, "Learner-Centered Education." 26.

have an alienated attitude about learning and believe that learning is “something you do in school, for school.”<sup>764</sup> He suggests that students:

are often motivated not by the desire for enhanced understanding and intellectual powers, but rather to satisfy (what they often see as arbitrary) requirements. We do not believe that any teacher sets out to induce such superficial learning in students, but all too often curricular traditions, pedagogy, or school structure encourage students to adopt just such an ‘alienated’ posture towards their learning.<sup>765</sup>

Even among students who do not feel alienated about learning, the use of “unnecessarily powerful extrinsic rewards” can counteract intrinsic motivation to learn.<sup>766</sup> This is especially true in situations where high-stakes or large-scale assessments are used. Shepard writes:

Established findings from the motivational literature have raised serious questions about test-based incentive systems. Students who are motivated by trying to do well on tests, instead of working to understand and master the material, are consistently disadvantaged in subsequent endeavors. They become less intrinsically motivated, they learn less, and they are less willing to persist with difficult problems.<sup>767</sup>

Performance assessments offer one way to counteract the connection between high-stakes assessment and low student motivation. According to Battersby, students are motivated by educational projects that “enable them to lead a richer and more empowered life rather than...a task done primarily to satisfy the demands of others (passing the test).”<sup>768</sup> Performance assessments

---

<sup>764</sup> Battersby, "So, What's a Learning Outcome Anyway?."

<sup>765</sup> Battersby, "So, What's a Learning Outcome Anyway?."

<sup>766</sup> Mark R. Lepper and Jennifer Henderlong, "Motivation: Instruction," Encyclopedia of Education, ed. James W. Guthrie, vol. 5 (New York: Macmillan Reference USA, 2003). 1696.

<sup>767</sup> Shepard, "Standardized Tests and High-Stakes Assessment." 2536.

<sup>768</sup> Battersby, "So, What's a Learning Outcome Anyway?."

fit this description. They provide “learner-centered, problem-driven approaches...[and] are most effective in engagement, motivation, and, through their problem-driven format, in providing a solid conceptual understanding.”<sup>769</sup> Performance assessment tasks can also focus on students’ interests and concerns, allowing teachers and assessors to support students’ values, interests, and relationships.<sup>770</sup> By using performance assessments, teachers may be able to encourage students’ intrinsic motivation and “foster mastery orientations.”<sup>771</sup> According to Lepper and Henderlong:

A classroom climate that supports mastery orientations—by minimizing public evaluation and normative comparisons, providing opportunities for improvement, and recognizing student effort—should also be beneficial. Instructional practices, therefore, should promote autonomy and minimize unnecessary extrinsic constraints, to foster intrinsic motivation and lifelong learning.<sup>772</sup>

Instructors who encourage intrinsic motivation allow students to have some control of their learning, and as a result, students feel a sense of competence and mastery.<sup>773</sup> According to Lepper and Henderlong, students who feel that their behavior is self-determined are motivated even when extrinsic rewards are absent.<sup>774</sup> They also point out studies demonstrating that individuals who feel in

---

<sup>769</sup> Norman and Spohrer, "Learner-Centered Education." 26-27.

<sup>770</sup> Lepper and Henderlong, "Motivation: Instruction." 1697.

<sup>771</sup> Lepper and Henderlong, "Motivation: Instruction." 1697.

<sup>772</sup> Lepper and Henderlong, "Motivation: Instruction." 1697.

<sup>773</sup> Small, Zakaria and El-Figuigui, "Motivational Aspects of Information Literacy Skills Instruction." 99.

<sup>774</sup> Lepper and Henderlong, "Motivation: Instruction." 1696.

control are more likely to participate in active learning, perceive greater competence, and achieve more.<sup>775</sup>

When students regulate their own learning, they exhibit greater motivation. According to Pintrich, self-regulated learning refers to “the processes by which individual learners attempt to monitor and control their own learning.”<sup>776</sup> Self-regulated learning is an “active, constructive process whereby learners set goals for their learning and then attempt to monitor, regulate, and control their cognition, motivation, and behavior, guided and constrained by their goals and the contextual features in the environment.”<sup>777</sup> Thus, good learners regulate their own motivation and emotions. According to Pintrich, self-regulation can be more significant than personal and environmental factors that impact learning; in fact, self-regulation processes allow students to “mediate the relations between the person, context, and eventual achievement.”<sup>778</sup>

Pintrich offers a number of generalizations about self-regulated learning and motivation. First, he notes that students must feel confident that they can complete a given task.<sup>779</sup> He writes, “Students who believe they can learn and are confident in their skills are more likely to report the use of self-regulatory

---

<sup>775</sup> Lepper and Henderlong, "Motivation: Instruction." 1696.

<sup>776</sup> Paul R. Pintrich, "Motivation: Self-Regulated Learning," Encyclopedia of Education, ed. James W. Guthrie, vol. 5 (New York: Macmillan Reference USA, 2003). 1698.

<sup>777</sup> Pintrich, "Motivation: Self-Regulated Learning." 1699.

<sup>778</sup> Pintrich, "Motivation: Self-Regulated Learning." 1699.

<sup>779</sup> Pintrich, "Motivation: Self-Regulated Learning." 1700.

strategies.”<sup>780</sup> Students must also value classroom tasks and be interested in them.<sup>781</sup> “Students who believe that their coursework is interesting, important and useful are more likely to report the use of self-regulatory strategies.”<sup>782</sup> Third, students who adopt a “mastery goal orientation” use self-regulatory strategies.<sup>783</sup> Pintrich believes, “If students set as their goal self-improvement and learning, then they will be much more likely to continue to engage in various cognitive and metacognitive activities in order to improve their learning and comprehension. The goal or criterion of learning and mastery seems to be a much better standard for self-regulated learning than an extrinsic goal.”<sup>784</sup> Thus, assessments that allow students to exhibit self-regulation are more closely aligned with motivation theories than are high-stakes, large-scale assessments that capitalize primarily on extrinsic rewards.

Brookhart developed a theoretical framework that attempts to explain the role of classroom assessment in motivating student achievement. In this framework, she merges together classroom assessment environment literature and educational motivation literature.<sup>785</sup> This framework (see Figure 3.6)

---

<sup>780</sup> Paul R. Pintrich, "The Role of Motivation in Promoting and Sustaining Self-Regulated Learning," International Journal of Education Research 31.6 (1999).

<sup>781</sup> Pintrich, "Motivation: Self-Regulated Learning." 1700.

<sup>782</sup> Pintrich, "The Role of Motivation."

<sup>783</sup> Pintrich, "The Role of Motivation."

<sup>784</sup> Pintrich, "The Role of Motivation."

<sup>785</sup> Susan M. Brookhart, "A Theoretical Framework for the Role of Classroom Assessment in Motivating Student Effort and Achievement," Applied Measurement in Education 10.2 (1997). 161.

supports the idea of improving and elevating the importance of classroom assessments.<sup>786</sup> It also reinforces the idea of using a variety of student

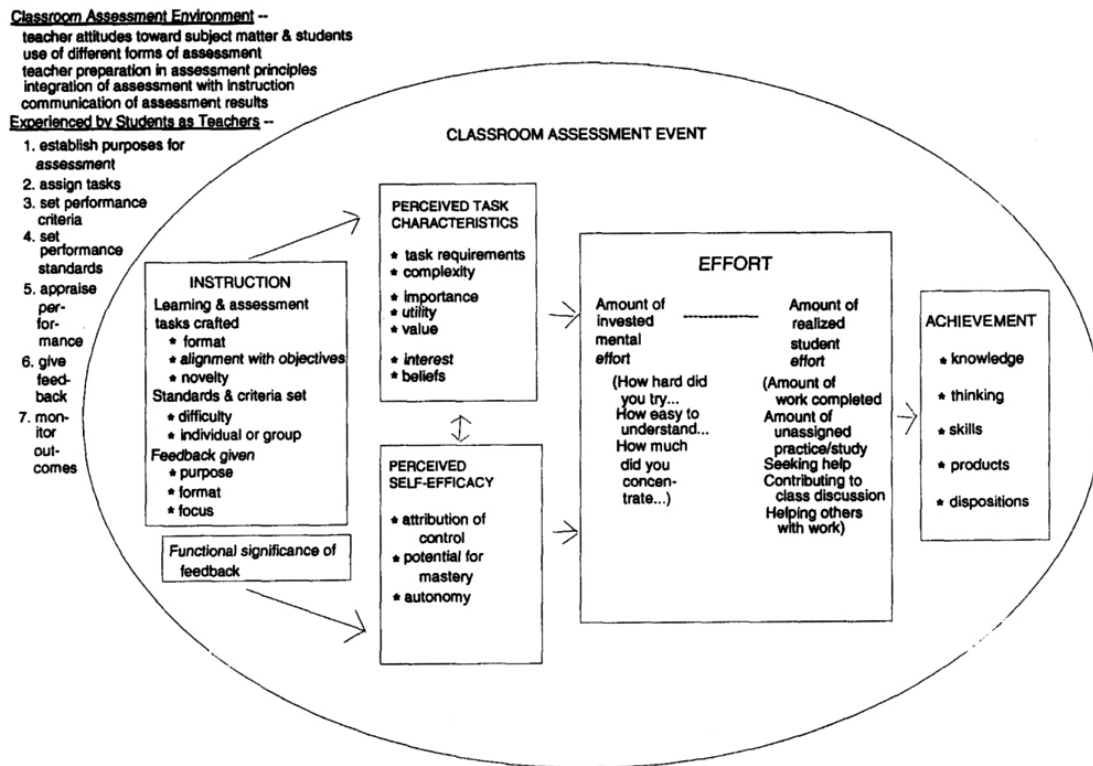


Figure 3.6. Assessment and Motivation

performances for assessment. According to Brookhart, "Assessment tasks help communicate the classroom assessment environment to the student and at the same time influence effort and achievement particularly through their perceived interests, utility, and relevance as goals."<sup>787</sup> Finally, Brookhart's theoretical framework emphasizes the importance of involving students in assessment processes. She writes:

<sup>786</sup> Brookhart, "A Theoretical Framework for the Role of Classroom Assessment." 178.

<sup>787</sup> Brookhart, "A Theoretical Framework for the Role of Classroom Assessment." 178.

Sharing ownership of assessment between teacher and students should enhance both the environmental and individual aspects of the process. A teacher who shares ownership of assessment communicates trust in students and confidence in their abilities to understand and apply performance criteria. A student who shares in the assessment process should perceive more control of and more responsibility for his or her own learning, which should increase effort and achievement.<sup>788</sup>

The assessment of information literacy instruction must incorporate these motivational theories to be successful. Baker and McKinzie humorously describe the lack of motivation that many teaching librarians encounter:

The kind of audience that most librarians face every day—peopled with conscripts, reluctant road warriors of the information highway. These students are a hardened lot, sophisticated yet ignorant, cynical yet strangely naïve. Most of them have little enthusiasm for a lecture of demonstration on the use of library sources. Even a spirited account of the wonders of the Internet or the mysteries of the online catalog can sometimes fail to engage them. Attending an instruction on research for most of them is just another of those things that one has to do.... Students regard instruction in research along with irritating roommates, political correctness, and even the college's cafeteria food as simply things that have to be endured—required (if not necessary) evils in the contemporary academy.<sup>789</sup>

Although many librarians recognize the students Baker and McKinzie describe, “to date, research on and development of [information literacy] skills instruction have focused exclusively on the content (the research process) or learning outcomes, with little or no attention paid to presentation methods that influence student motivation.”<sup>790</sup> Yet, according to Kilcullen, “Knowing what motivates students...is essential to teaching critical thinking skills in library users,” and

---

<sup>788</sup> Brookhart, "A Theoretical Framework for the Role of Classroom Assessment." 178.

<sup>789</sup> Neal Baker and Steve McKinzie, "Librarians and Faculty in Tandem: Taking Our Cue from the Evening News," *Reference and User Services Quarterly* 37.1 (1997).

<sup>790</sup> Small, Zakaria and El-Figuigui, "Motivational Aspects of Information Literacy Skills Instruction." 98.

considering student motivation will help librarians become better teachers.<sup>791</sup> Indeed, students put the responsibility to create motivating rather than boring lessons squarely on the shoulders of their instructors.<sup>792</sup> Kasowitz-Schneer and Pasqualoni recognize that “one of the major challenges facing the [information literacy] instructor is to provide the type of...experiences that motivate students to learn and apply...skills.”<sup>793</sup> Few authors to tackle the issue of motivation in information literacy instruction in higher education; however, Baker and McKinzie suggest that teaching librarians collaborate with course instructors to learn about motivation techniques.<sup>794</sup> Kasowitz-Schneer and Pasqualoni list “motivating students to learn information literacy skills” as a challenge of information literacy instruction.<sup>795</sup> Certainly there is more work to be done, not only to connect motivation theory to information literacy instruction, but also to the assessment of that instruction.

### ***Theoretical Background: Assessment***

In addition to educational and motivational theories, outcomes-based performance assessment is grounded in “assessment for learning” theories.

---

<sup>791</sup> Maureen Kilcullen, "Teaching Librarians to Teach: Recommendations on What We Need to Know," Reference Services Review 26.2 (1998). 9.

<sup>792</sup> Ruth V. Small, B. M. Dodge and X. Jiang, "Dimensions of Interest and Boredom in Instructional Situations," 1996 National Convention of the Association for Educational Communications and Technology (Indianapolis, Indiana: 1996).

<sup>793</sup> Abby Kasowitz-Scheer and Michael Pasqualoni, "Information Literacy Instruction in Higher Education: Trends and Issues," ERIC Digest ED465375 (2002).

<sup>794</sup> Baker and McKinzie, "Librarians and Faculty in Tandem."

<sup>795</sup> Kasowitz-Scheer and Pasqualoni, "Information Literacy Instruction in Higher Education."



Assessment for learning theory suggests that “good teaching is inseparable from good assessing.”<sup>796</sup> Assessment for learning theorists believe the connection between teaching and testing can “lead to a substantial increase in instructional effectiveness.”<sup>797</sup> Battersby argues that assessment should be thought of not just as evaluation, but as a “primary means” of learning that requires the use of meaningful and complex assessment assignments.<sup>798</sup> He writes, “Assessment is directly related to learning, not merely to evaluation or certification.”<sup>799</sup> Arter agrees that assessments can be tools for learning, and that students should learn by completing an assessment.<sup>800</sup> She explains, “Educators do not teach and then assess; nor do they think of assessment as something that is done to students. Instead, they consider the assessment activity itself an instructional episode.”<sup>801</sup>

When an assessment is elevated to a learning tool, assessors should aspire to additional goals. First, performance assessments should be based on the curriculum so that the curriculum, rather than non-local expectations, “drives” the testing activities.<sup>802</sup> Second, performance assessments should be “worth

---

<sup>796</sup> Wiggins, "Creating Tests Worth Taking." V-6: 8.

<sup>797</sup> Popham, Test Better, Teach Better. 1.

<sup>798</sup> Battersby, So, What's a Learning Outcome Anyway? Learning Outcomes and the Learning Paradigm.

<sup>799</sup> Battersby, So, What's a Learning Outcome Anyway? Learning Outcomes and the Learning Paradigm.

<sup>800</sup> Arter, "Using Assessment." IV-10: 1.

<sup>801</sup> Arter, "Using Assessment." IV-10:1.

<sup>802</sup> Sweet, "Performance Assessment."

teaching to” and present interesting, meaningful, and complex tasks to students.<sup>803</sup> Sweet notes that they should allow students to bring a variety of curriculum-related skills to bear.<sup>804</sup> Ideally, the assessment task should be one that is already inherent in the curriculum or learning program.<sup>805</sup> If an embedded task cannot be found, then the assessment task should be one that could be easily included in day-to-day instruction.<sup>806</sup>

According to Shepard, assessment for learning should allow students to exhibit hallmarks of Glasser’s competence indicators: coherence of knowledge, principled problem solving, knowledge use, automatized skills, and metacognitive or self-regulatory skills.<sup>807</sup> To do this, she recommends using assessments that provide opportunities for students to demonstrate their ability to connect concepts, recognize underlying themes, principles, and patterns important for problem solving, and monitor their own learning.<sup>808</sup> Assessment tasks that offer all of these qualities, according to Shepard, are easily found among regular classroom learning tasks.<sup>809</sup>

Marzano, Pickering, and McTigue have developed a framework for conceptualizing “dimensions of learning” that suggests linkages between

---

<sup>803</sup> Sweet, "Performance Assessment."

<sup>804</sup> Sweet, "Performance Assessment."

<sup>805</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 22.

<sup>806</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 22.

<sup>807</sup> Shepard, "Why We Need Better Assessments." I-2: 4.

<sup>808</sup> Shepard, "Why We Need Better Assessments." I-2: 4.

<sup>809</sup> Shepard, "Why We Need Better Assessments." I-2: 4.

instruction and performance-based assessments. The framework helps clarify the connections between performance assessment and learning by outlining assumptions about their shared nature.<sup>810</sup> The dimensions of learning framework includes five kinds of thinking that are required for successful learning (see Figure 3.7). The framework shows that successful learners exhibit positive attitudes and perceptions, integrate new knowledge with old, extend and refine their knowledge, use knowledge to perform meaningful tasks, and develop productive habits of mind.<sup>811</sup> These five dimensions work together when students engage in complex learning and assessment tasks that require them to use knowledge to make decisions, investigate, experiment, solve problems, and invent.<sup>812</sup> Marzano, Pickering, and McTighe lament that many traditional assessments attempt none of these things, but hold out hope that outcomes-based performance assessments will continue to be linked to “the kind of learning we want to see.”<sup>813</sup>

In the realm of information literacy instruction, outcomes-based performance assessments that emphasize “assessment for learning” are slow to take shape. Yet, Warner offers hope, suggesting that “building the assessment process into the entire library instruction program is developing naturally.”<sup>814</sup>

---

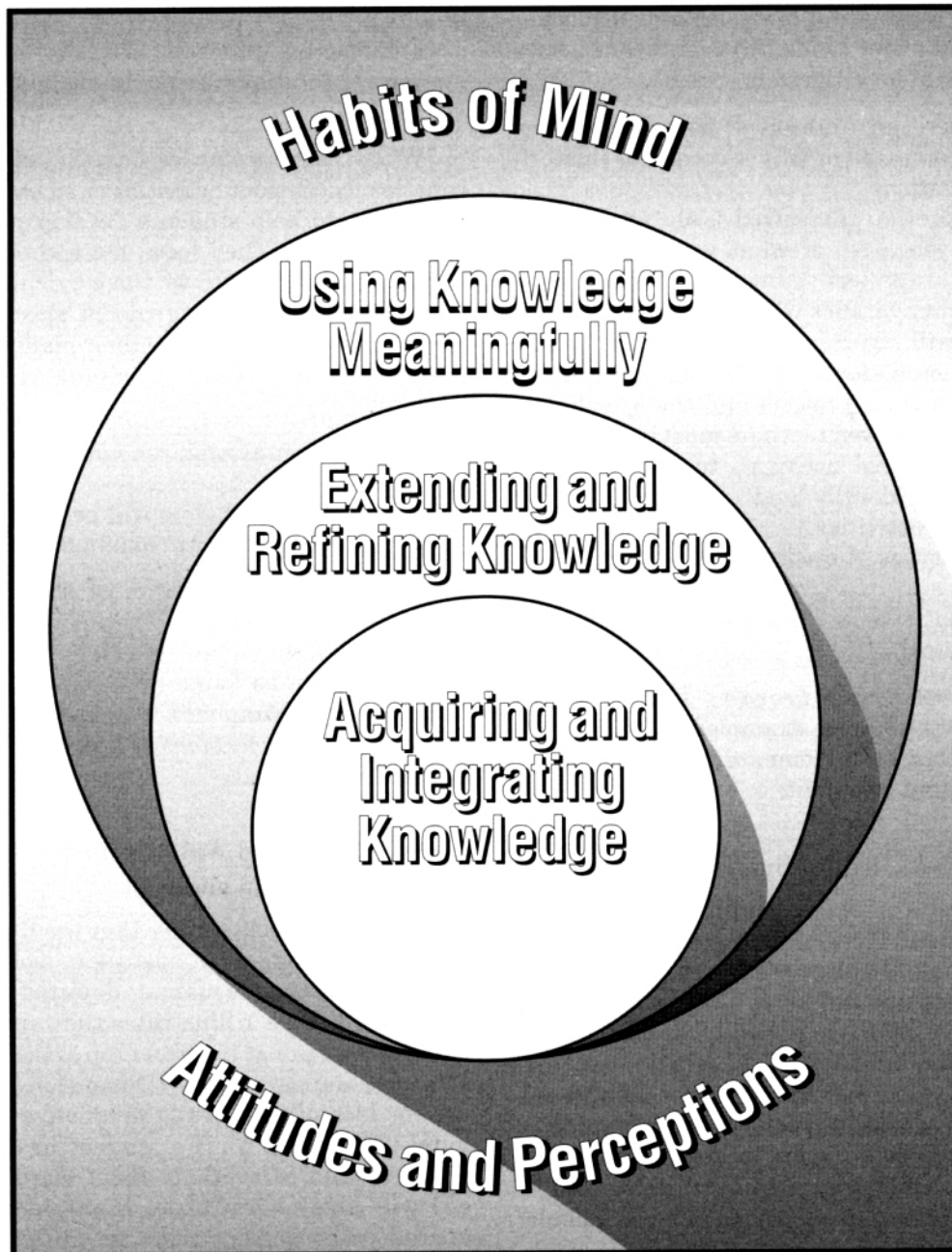
<sup>810</sup> Marzano, Pickering and McTighe, Assessing Student Outcomes. 1.

<sup>811</sup> Marzano, Pickering and McTighe, Assessing Student Outcomes. 1-3.

<sup>812</sup> Marzano, Pickering and McTighe, Assessing Student Outcomes. 3.

<sup>813</sup> Marzano, Pickering and McTighe, Assessing Student Outcomes. 5.

<sup>814</sup> Warner, "Programmatic Assessment." 175.



*Figure 3.7. Thinking Required for Successful Learning*

Grassian and Kaplowitz also describe the potential for assessment for learning in information literacy instruction:

Our learners can also gain from the assessment process. As they reflect on the instruction, what they have learned, and how that information has been useful to them, learners begin to explore the learning process itself, thus engaging in the metacognition process. They delve into how they

interacted with the information being presented and consider how they might do this more effectively in the future. A well-designed assessment not only provides useful information for the instructor, it actually benefits the learner and helps to reinforce the material that was taught. Research has indicated that people who become aware of themselves as learners—that is, those who are self-reflective and analytic about their own learning process—become better learners. They move from being ‘surface learners’ who merely reproduce information provided by others to ‘deep learners’ who not only understand the information, but can apply it appropriately in a variety of settings (Corno and Mandinach, 1983; Cross, 1998). As a result, thoughtfully designed assessments can enhance the students’ abilities to become life-long learners. Assessment, therefore, contributes to the overall goals of ILI. It enhances the learners’ experience by allowing them to examine how they learn and to develop more efficient and effective IL strategies and skills.<sup>815</sup>

Hopefully, the use of assessment for learning strategies used in information literacy instruction will increase in the coming years.

### ***Benefits***

The benefits of outcomes-based performance assessments are so great that Silver suggests that “teaching to the test”, a strategy that is usually listed as a disadvantage of assessment approaches, is actually an advantage of performance assessment.<sup>816</sup> Performance assessment offers numerous benefits to teachers and learners. Among them are the close connections between instruction and assessment, the ability to measure higher-order thinking skills, the contextualization of assessment that leads to greater equitability and validity, and the ability to use results to improve instruction and programs.

---

<sup>815</sup> Grassian and Kaplowitz, Information Literacy Instruction. 287.

<sup>816</sup> Silver, "Performance Assessment." 135.

Perhaps the greatest value of performance assessment is that the form and content of the assessment method can be closely aligned with instructional goals.<sup>817</sup> As a result of this alignment, “the use of performance assessment in the classroom has been seen by some as a promising means of accomplishing a long-standing, elusive goal—namely, the integration of instruction and assessment.”<sup>818</sup> Because instruction and assessment are integrated in performance assessments, educators can learn about a broader range of learning outcomes and, at the same time, “preserve the complex nature of disciplinary knowledge and inquiry, including conceptual understanding, problem-solving skills, and the application of knowledge and understanding to unique situations.”<sup>819</sup>

Performance assessments allow educators to capture students’ learning of higher-order thinking and reasoning skills, skills that are typically absent in more traditional forms of assessment.<sup>820</sup> Resnik describes the ways in which performance task focus on higher-order skills. For example, she explains that higher-order thinking is nonalgorithmic, because the path of action is not preconceived or provided to a student.<sup>821</sup> She states that higher-order thinking is complex in that the “total path of action is not ‘visible’ (mentally speaking) from

---

<sup>817</sup> Silver, "Performance Assessment." 135.

<sup>818</sup> Silver, "Performance Assessment." 135.

<sup>819</sup> Silver, "Performance Assessment." 135.

<sup>820</sup> Silver, "Performance Assessment." 135.

<sup>821</sup> Lauren B. Resnick, Education and Learning to Think (Washington, D.C.: National Academy Press, 1987). 3.

any single vantage point.”<sup>822</sup> Such thinking yields multiple solutions and requires nuanced judgment.<sup>823</sup> Students using higher-order skills have to apply multiple, possibly conflicting criteria, deal with uncertainty and the unknown, and to do so they must regulate their own thinking.<sup>824</sup> Finally, they must impose meaning on structures that may initially seem meaningless,<sup>825</sup> which takes a great deal of effort and mental work.<sup>826</sup> All of these characteristics of higher-order thinking are present in outcomes-based performance assessments.

Another benefit of outcomes-based performance assessments is that they are contextualized. According to Wiggins, “There is no such thing as performance-in-general. To understand what kind and precision of answer fits the problem at hand, the student needs contextual detail.”<sup>827</sup> Performance assessments recognize the contexts in which students work and aim to “invent an authentic simulation...and like all simulations...the task must be rich in contextual detail.”<sup>828</sup> Through contextualization, performance assessments help students understand the relevance of what they learn.<sup>829</sup> Performance assessments also can “reflect...society’s demands for education that prepares

---

<sup>822</sup> Resnick, Education and Learning to Think. 3.

<sup>823</sup> Resnick, Education and Learning to Think. 3.

<sup>824</sup> Resnick, Education and Learning to Think. 3.

<sup>825</sup> Resnick, Education and Learning to Think. 3.

<sup>826</sup> Resnick, Education and Learning to Think. 3.

<sup>827</sup> Wiggins, "Creating Tests Worth Taking." V-6: 5.

<sup>828</sup> Wiggins, "Creating Tests Worth Taking." V-6: 3.

<sup>829</sup> Farmer, "Authentic Assessment of Information Literacy."

students for the world of work.”<sup>830</sup> According to Farmer, “Authentic assessment helps bridge the two worlds.... They show that [students] can apply theoretical concepts to solving life-like problems. As a result, education doesn’t seem to operate in a vacuum; it truly prepares students for the rest of their lives. [It is] real learning for real results.”<sup>831</sup>

Through the contextualization of performance assessment, three more benefits are realized: transfer of knowledge, equitability, and validity. Shepard states that good teaching and good assessment “ask...about old understandings in new ways, call...for new applications, and draw...new connections.”<sup>832</sup> She argues that by contextualizing assessment in real world problems, students demonstrate their ability transfer knowledge and use it in new ways.<sup>833</sup> Using contextualized, authentic problems in assessment is also a more equitable approach, as opposed to using tests that are susceptible to the “bias associated with testing rapid recall of decontextualized information.”<sup>834</sup> Finally, performance assessments may be able to render more valid data than other types of assessments. Sweet suggests that “because they require students to actively demonstrate what they know, performance assessments may be a more valid indicator of student’s knowledge and abilities. There is a big difference between answering multiple-choice questions on how to make an oral presentation and

---

<sup>830</sup> Farmer, "Authentic Assessment of Information Literacy."

<sup>831</sup> Farmer, "Authentic Assessment of Information Literacy."

<sup>832</sup> Shepard, "The Role of Assessment in a Learning Culture." 11.

<sup>833</sup> Shepard, "The Role of Assessment in a Learning Culture." 11.

<sup>834</sup> Silver, "Performance Assessment." 135.



actually making an oral presentation.”<sup>835</sup> Silver also notes that “performance measures offer a potential advantage of increased validity over other forms of testing that rely on indirect indicators of a desired competence or proficiency.”<sup>836</sup>

Outcomes-based performance assessments offer additional programmatic benefits as well. Sweet states that performance assessment “can provide impetus for improving instruction.”<sup>837</sup> This type of assessment is suitable for focusing on the goals of an instructional program and allows documentation of how students are developing and learning.<sup>838</sup> Such documentation gives program administrators data with which they can improve the program or argue for greater resources.<sup>839</sup> Other authors acknowledge that the documentation resulting from performance assessments is suitable for communicating issues and needs to constituents. They write, “Policy makers at all levels have embraced performance assessments as outcomes measures suitable for reporting to parents and legislative bodies.”<sup>840</sup>

### ***Limitations***

According to Silver, there are a few technical and feasibility issues that have, in the past, thwarted attempts to use performance assessment on a large

---

<sup>835</sup> Sweet, "Performance Assessment."

<sup>836</sup> Silver, "Performance Assessment."

<sup>837</sup> Sweet, "Performance Assessment."

<sup>838</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 22.

<sup>839</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 22.

<sup>840</sup> Baker, Abedi, Linn and Niemi, "Dimensionality and Generalizability." 197.

scale.<sup>841</sup> Among these are cost, time, and generalizability. For example, compared to traditional tests, performance assessments can be costly to develop, administer, and score.<sup>842</sup> However, Silver offers hope that, in the future, new technologies will decrease the costs associated with test development, administration, and scoring.<sup>843</sup> Another limitation of an outcomes-based performance assessment approach is time. Sweet notes that performance assessments require greater time, in both planning and thought, from both teachers and students.<sup>844</sup> Popham adds that teachers spend more time because they design multiple steps of a performance, create a rubric, and use the rubric to score the performance.<sup>845</sup> More thought also must be given to the outcomes to be assessed. Bresciani, Zelna, and Anderson remind the creators of performance assessments that "writing measurable and meaningful outcomes is typically not an overnight assignment."<sup>846</sup> Baker also points out that the need to carefully consider outcomes increases as the stakes of a performance assessment increase.<sup>847</sup>

---

<sup>841</sup> Silver, "Performance Assessment." 135.

<sup>842</sup> Silver, "Performance Assessment." 135.

<sup>843</sup> Silver, "Performance Assessment." 136.

<sup>844</sup> Sweet, "Performance Assessment."

<sup>845</sup> Popham, Test Better, Teach Better. 102.

<sup>846</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 22.

<sup>847</sup> Baker, Abedi, Linn and Niemi, "Dimensionality and Generalizability." 197.

A third limitation of outcomes-based performance assessment approaches is the generalizability and comparability of results.<sup>848</sup> Popham asks, “How many performance tests do students need to complete before the teacher can come up with valid inferences about their generalizable skill-mastery?”<sup>849</sup> He fears that many tests may be needed and, as a result, cautions educators to use performance assessment to assess “only the most significant of your high-priority curricular aims.”<sup>850</sup> Still, this concern about the generalizability of performance assessments may be only temporary. According to Silver, “Generalizability across tasks may be increased through the use of intelligent systems that offer ongoing assessment well integrated with instruction and sensitive to changes in students’ understanding and performance, with performance data collected over a long period of time as opposed to one-time, on-demand testing.”<sup>851</sup> Overall, Silver suggests that advances in cognitive sciences and technology may resolve the limitations of performance assessment.<sup>852</sup>

### ***Large-Scale Accountability Testing and Classroom Assessment***

Based on the benefits and limitations of outcomes-based performance assessment, an argument could be made that while performance assessment has many advantages over traditional surveys and tests, its limitations make it

---

<sup>848</sup> Silver, “Performance Assessment.” 135.

<sup>849</sup> Popham, Test Better, Teach Better. 102.

<sup>850</sup> Popham, Test Better, Teach Better. 103.

<sup>851</sup> Silver, “Performance Assessment.” 136.

<sup>852</sup> Silver, “Performance Assessment.” 136.

difficult or impossible to adapt to large-scale testing situations. Certainly, in the past, large-scale accountability testing and performance-focused classroom assessment have had very different characteristics. For example, large-scale assessments are conceptualized as “formal, objective, time efficient, cost-efficient, widely applicable, and centrally processed.”<sup>853</sup> Above all, they must be useful to decision makers, which may be interpreted as being reduced to a single score.<sup>854</sup> Traditional large-scale assessments are typically not useful in the “diagnosis and targeting of individual student learning needs.”<sup>855</sup> This is widely recognized, and Hoyler notes, “even as we use [large-scale tests], there is little indication that we place undue faith in them.”<sup>856</sup> Even so, Shepard suggests that the goal “should be to find ways to fend off the negative effects of externally imposed tests and to develop instead classroom assessment practices that can be trusted to help students take the next steps in learning.”<sup>857</sup>

In contrast, classroom assessments are designed to support instruction. They are usually more informal, and they are locally developed, scored, and interpreted.<sup>858</sup> They are more likely to include tasks that are instructionally

---

<sup>853</sup> Shepard, "Why We Need Better Assessments." I-2: 5.

<sup>854</sup> Shepard, "Why We Need Better Assessments." I-2: 5.

<sup>855</sup> Shepard, "Standardized Tests and High-Stakes Assessment." 2534.

<sup>856</sup> Hoyler, "The Road Not Taken."

<sup>857</sup> Shepard, "The Role of Assessment in a Learning Culture." 12.

<sup>858</sup> Shepard, "Why We Need Better Assessments." I-2: 5.

valuable, they are capable of showing short-term changes in student learning, and they provide meaningful feedback to students.<sup>859</sup> Shepard explains:

In classrooms, assessment is an integral part of the teaching and learning process. Teachers use both formal and informal assessments to plan and guide instruction. For individual students, assessments help to gauge what things students already know and understand, where misconceptions exist, what skills need more practice in context, and what supports are needed to take the next steps in learning. Teachers also use assessment to evaluate their own teaching practices so as to adjust and modify curricula, instructional activities, or assignments that did not help students grasp key ideas. To serve classroom purposes, assessments must be closely aligned with what children are learning, and the timing of assessments must correspond to the specific days and weeks when children are learning specific concepts.<sup>860</sup>

Not only can the timing of classroom assessments be better than large-scale testing approaches, the sampling of student responses can also be more effective. Stiggins acknowledges that the perfect sample of student performance is elusive.<sup>861</sup> Still, he notes that classroom teachers must compromise on student samples less than large-scale assessors because they have more time with students.<sup>862</sup> For this reason, he states, "the great strength and future of performance assessment lies in the classroom, not in large-scale standardized testing."<sup>863</sup> A final advantage of classroom assessment is the lack of political pitfalls. Because of the strong links between instruction and assessment,

---

<sup>859</sup> Shepard, "Why We Need Better Assessments." I-2: 5.

<sup>860</sup> Shepard, "Standardized Tests and High-Stakes Assessment." 2534.

<sup>861</sup> Richard J. Stiggins, "The Development of Performance Assessment Exercises," A Handbook for Student Performance in an Era of Restructuring, eds. Robert E. Blum and Judith A. Arter (Alexandria, Virginia: Association for Supervision and Curriculum Development, 1996). V-8: 7.

<sup>862</sup> Stiggins, "The Development of Performance Assessment Exercises." V-8: 7.

<sup>863</sup> Stiggins, "The Development of Performance Assessment Exercises." V-8: 7.

classroom assessments are “conducted in a climate of greater trust” than large-scale assessments.<sup>864</sup> Thus, the lack of political pressure allows teachers to assess a wider range and a higher level of learning outcomes.<sup>865</sup>

Despite the many differences between these approaches, there may be ways to combine large-scale testing and classroom assessment to provide decision-makers with necessary data and support learning in the classroom at the same time. Evans states, “There are times when the same assessment measures can be used to provide important data for change to occur either at the institution, the program, or the classroom level or even at all three levels.”<sup>866</sup>

Arter believes that “the whole enterprise [of assessment] is much more powerful...if large-scale and classroom assessment are considered as a unit.”<sup>867</sup>

Arter suggests that this could be accomplished if performance criteria used in large-scale assessments were developed by teachers with classroom uses in mind and if these performance criteria were uniform across teachers, so that “we could be confident that, regardless of who observes a student performance, the judgment would be the same.”<sup>868</sup>

---

<sup>864</sup> Shepard, “Why We Need Better Assessments.” I-2: 5.

<sup>865</sup> Shepard, “Why We Need Better Assessments.” I-2: 5.

<sup>866</sup> Evans, “Understanding Assessment.” 8.

<sup>867</sup> Arter, “Using Assessment.” IV-10: 5.

<sup>868</sup> Arter, “Using Assessment.” IV-10: 5.

### ***Performance Assessments of Information Literacy Skills***

Although surveys and tests far outnumber performance assessments of information literacy skills, a few significant examples are worth noting. For example, Young and Ackerson,<sup>869</sup> Hovde,<sup>870</sup> Flaspohler,<sup>871</sup> and Ursin, Lindsay, and Johnson<sup>872</sup> describe assessments based on the analysis of bibliographies and paper citations. Snaveley and Wright<sup>873</sup> and Fourie and Van Niekerk<sup>874</sup> report their experiences with portfolio assessment. Horan describes the use of sketch maps in a creative and different approach to information literacy assessment.<sup>875</sup>

In addition to these performance assessments, the Educational Testing Service (ETS) has recently developed a performance-based test of information and communication technology (ICT) skills. This test was born out of a movement to address and reduce the digital divide on a global level by

---

<sup>869</sup> Virginia E. Young and Linda G. Ackerson, "Evaluation of Student Research Paper Bibliographies: Refining Evaluation Criteria," Research Strategies 13.2 (1995).

<sup>870</sup> Karen Hovde, "Check the Citation: Library Instruction and Student Paper Bibliographies," Research Strategies 17 (2000).

<sup>871</sup> Flaspohler, "Information Literacy Program Assessment."

<sup>872</sup> Lara Ursin, Elizabeth Blakesley Lindsay and Corey M. Johnson, "Assessing Library Instruction in the Freshman Seminar: A Citation Analysis Study," Reference Services Review 32.3 (2004).

<sup>873</sup> Loanne L. Snaveley and Carol A. Wright, "Research Portfolio Use in Undergraduate Honors Education: Assessment Tool and Model for Future Work," Journal of Academic Librarianship 29.5 (2003).

<sup>874</sup> Ina Fourie and Daleen Van Niekerk, "Using Portfolio Assessment in a Module in Research Information Skills," Education for Information 17.4 (1999).

<sup>875</sup> Mark Horan, "What Students See: Sketch Maps as Tools for Assessing Knowledge of Libraries," Journal of Academic Librarianship 25.3 (1999).

examining not just access to technology, but also the ability to use and integrate technology into life and work.<sup>876</sup> The International ICT Literacy Panel states:

As technology approaches ubiquity, an increasing importance must now be placed on educating and training citizenry in the ICT skills necessary to function effectively in a global economy increasingly dependent on ICT. The panel's overarching belief is that the digital divide should no longer be defined only in terms of limited access to hardware, software, and networks, but rather one that is also driven by limited literacy skills and a lack of the cognitive skills needed to make effective use of these technologies. Technology skills alone, without corresponding cognitive skills and general literacy, will not decrease the gaps defined by a digital divide.<sup>877</sup>

These 'corresponding cognitive skills and general literacy' skills include reading, numeracy, and problem-solving skills.<sup>878</sup> The panel recommended gathering "meaningful data from large scale global assessments, and from smaller diagnostic tests aimed to inform governments, schools, and private sector organizations and consortiums" in order to describe and gauge the gaps in ability to use information and communication technologies.<sup>879</sup>

ETS defines ITC literacy as "using digital technology, communication tools, and/or networks to access, manage, integrate, evaluate, and create information in order to function in a knowledge society."<sup>880</sup> ETS developed a framework of ICT literacy (see Figure 3.8). In this framework, cognitive

---

<sup>876</sup> International ICT Literacy Panel, Digital Transformation: A Framework for ICT Literacy, 2002, Educational Testing Service 2004 <<http://www.ets.org/research/ictliteracy/>>. 6-7.

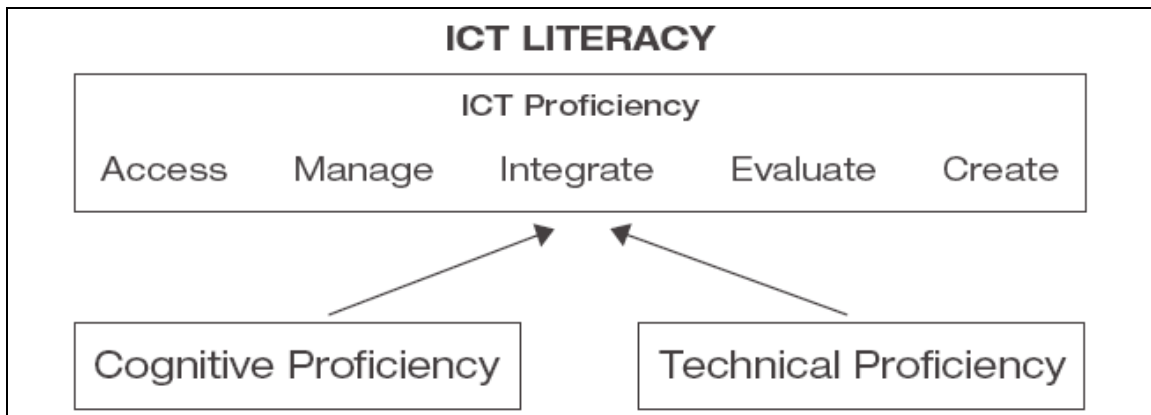
<sup>877</sup> International ICT Literacy Panel, Digital Transformation.

<sup>878</sup> International ICT Literacy Panel, Digital Transformation. 1.

<sup>879</sup> International ICT Literacy Panel, Digital Transformation. 1.

<sup>880</sup> International ICT Literacy Panel, Digital Transformation.





*Figure 3.8. ICT Literacy Framework*

proficiency refers to “the desired foundational skills of everyday life at school, at home, and at work. Literacy, numeracy, problem solving, and spatial/visual literacy demonstrate these proficiencies.”<sup>881</sup> Technical proficiency includes the “basic components of digital literacy. It includes a foundational knowledge of hardware, software applications, networks, and elements of digital technology.”<sup>882</sup> Through this framework, ETS attempts to convey that ICT literacy is more than just technology literacy, and that it includes critical and cognitive skills.<sup>883</sup>

Several limitations may impact the use of the ICT test to assess library instruction. One limitation is that the ICT test may not serve local needs. Librarians need to map ICT test tasks to the ACRL standards and determine whether the test assesses the learning outcomes of local library instruction. There may also be a disconnect between the goals of the ICT test and the

<sup>881</sup> International ICT Literacy Panel, Digital Transformation. 18.

<sup>882</sup> International ICT Literacy Panel, Digital Transformation. 18.

<sup>883</sup> International ICT Literacy Panel, Digital Transformation. 1.

teaching and learning missions of specific educational institutions. The ICT framework is intended to serve as the “basis for the design and conduct of large-scale national and international assessments as well as diagnostic tests of individual skills associated with information and communication technology.”<sup>884</sup> ETS lists ICT stakeholders of such assessments in this order: “government policy makers, corporate leaders, industry associations, unions, workforce groups, educators (K-12, higher education, national education associations, researchers), consumer and public interest groups, and relevant international associations.”<sup>885</sup> Because the ICT test has so many stakeholders, librarians need to determine whether or not the test meets their local need for assessment of library instruction.

Another limitation of the ICT test is that it appears to use score-spread techniques to differentiate the levels of student performance. In creating the test framework, the International ICT Literacy Panel focused on the variables that “account for large percentages of variation in the distribution of tasks.”<sup>886</sup> This approach leaves any assessments that grow from the framework open to the disadvantages of score-spread test techniques, which are outlined earlier in this chapter.

Limitations aside, the first test developed by ETS from the ITC literacy framework, rolled out in the spring of 2005, is attracting attention. One of the

---

<sup>884</sup> International ICT Literacy Panel, Digital Transformation. 3.

<sup>885</sup> International ICT Literacy Panel, Digital Transformation. 22.

<sup>886</sup> International ICT Literacy Panel, Digital Transformation.

main reasons is the performance focus of the test. According to the International ICT Literacy panel, all other tests of literacy, numeracy, problem solving, and technical knowledge are paper and pencil, rather than computer-based, tests.<sup>887</sup> According to ETS, Rockman stated that she compared other available testing options, and “there is none more comprehensive or distinctive than this performance-based, web-based, holistic assessment tool.”<sup>888</sup> Eschewing multiple-choice item types, the two-hour test asks students to respond to sixteen problem scenarios.<sup>889</sup> The content of these scenarios is described in different ways. One article states that the test will “measure students’ ability to evaluate online material.”<sup>890</sup> Another says that the test “is intended to measure students’ ability to manage exercises like sorting e-mail messages or manipulating tables and charts.”<sup>891</sup> Yet another states that it “will evaluate how well students can...build a spreadsheet, compose email messages summarizing a passage, and perform other tasks.”<sup>892</sup> A final article reports that the test requires “opening messages and attachments and copying relevant information; Internet search skills including creating search terms and effectively weeding through search

---

<sup>887</sup> International ICT Literacy Panel, Digital Transformation.

<sup>888</sup> Educational Testing Service, ETS Collaborates with Major Universities to Assess 21st Century Skills, 2003, March 21, 2004 <<http://www.ets.org/news/03091701.html>>.

<sup>889</sup> Kristin Bagnato, "New Exam Tests Students' IT Abilities," Community College Week (2005).

<sup>890</sup> Geoffrey Nunberg, "Teaching Students to Swim in the Online Sea," New York Times February 13 2005.

<sup>891</sup> Tom Zeller Jr., "Measuring Literacy in a World Gone Digital," New York Times January 17 2005.

<sup>892</sup> Jeffrey R. Young, "Testing Service to Unveil an Assessment of Computer and Information Literacy," Chronicle of Higher Education November 12 2004.

results; and word processing skills such as creating files and cutting and pasting information.”<sup>893</sup>

According to the ETS, “Unlike traditional assessments—which use discrete, artificial tasks to evaluate performance—the...assessments will evaluate ICT proficiency using complex tasks that simulate real-life demands.”<sup>894</sup> However, Asaravala feels that ETS may overstate this strength. He writes, “Test-takers are stuck in the simulated computer environment created by ETS. Forget about using your own e-mail client or the Mozilla web browser. And don’t even think about reaching for any of the dozens of keyboard shortcuts that you’ve grown accustomed to.”<sup>895</sup> Asaravala also found problems with the context of many of the tasks, and suggests that depending on a test-takers background, he or she may be at a disadvantage.<sup>896</sup> Asaravala quotes a student who took the test and thought that both the context and the wording of some tasks made them unnecessarily difficult. The student cited business situations she wasn’t familiar with, saying, “I looked at some of them for a long time, thinking, ‘What are they talking about?’...I think they have to work on [the test].”<sup>897</sup> In summary, these concerns point to important limitations of the ICT test tasks: they simulate real-

---

<sup>893</sup> Bagnato, "New Exam."

<sup>894</sup> Educational Testing Service, Succeeding in the 21st Century: What Higher Education Must Do to Address the Gap in Information and Communication Technology Proficiencies, 20032004 <<http://www.calstate.edu/LS/ICTwhitepaperfinal.pdf>>.

<sup>895</sup> Amit Asaravala, "Testing Your Tech Smarts," Wired News April 8 2005.

<sup>896</sup> Asaravala, "Testing Your Tech Smarts."

<sup>897</sup> Asaravala, "Testing Your Tech Smarts."

world tasks, but they are not contextualized or naturally-occurring in some students' school curriculum.

In the future, the ETS suggests that the ICT test could be used for a variety of purposes. According to ETS, the test is intended to "provide aggregated results for measuring the performance of particular groups...[so that] higher education administrators and faculty can determine and describe the ICT strengths and weaknesses" of students.<sup>898</sup> The results would provide placement or remediation indicators or could be used to benchmark institutional performance.<sup>899</sup> According to ETS, the test could "provide individual results that can be used to measure the basic ICT proficiency of a student or potential employee,"<sup>900</sup> or the test could be used as an admissions tool or a work-force certification tool.<sup>901</sup> It could also be used for resource allocation and curriculum planning, or to allow students to place out of courses.<sup>902</sup> During 2005 and 2006, ETS will provide aggregated results, but not individualized results.<sup>903</sup> Later on, scores reports will focus on individual students and consist of "strong", "satisfactory", or "poor" performance ratings.<sup>904</sup> These scores may or may not allow educators to make decisions for instructional improvement.

---

<sup>898</sup> Educational Testing Service, ICT Literacy Assessment.

<sup>899</sup> Educational Testing Service, ICT Literacy Assessment.

<sup>900</sup> Educational Testing Service, ICT Literacy Assessment.

<sup>901</sup> Bagnato, "New Exam."

<sup>902</sup> Young, "Testing Service."

<sup>903</sup> Young, "Testing Service."

<sup>904</sup> Young, "Testing Service."

Overall, response to the ICT test has been mixed. Some students seem confused by the content of the test.<sup>905</sup> Other observers question ETS's definition of information and technological literacy, and argue that "there is no widespread agreement on whether such skills can be taught, much less measured in a test."<sup>906</sup> Zeller suggests that ETS is showing a certain level of opportunism. Zeller writes, "What seems certain...is that a lucrative market is emerging for testing companies that are willing to fill the perceived need."<sup>907</sup> Alliance for Childhood states, "For ETS, this is part of a broader global plan to develop and promote international technology literacy standards, and then offer countries around the world a chance to buy a full array of assessment products and services that can be used to implement their standards."<sup>908</sup> After taking the IC test, Asaravala comments:

Whatever my score, the question remains: Does the ICT literacy assessment really measure a person's ability to make critical judgments and solve problems in today's tech-oriented world? It's possible that no one will know for sure until researchers have had time to follow test-takers to see whether high scorers really end up doing better in school and in the work force. Of course, by then information and communication technologies might look completely different than they do today, and ETS may find itself designing a whole new test..<sup>909</sup>

---

<sup>905</sup> Bagnato, "New Exam."

<sup>906</sup> Zeller Jr., "Measuring Literacy."

<sup>907</sup> Zeller Jr., "Measuring Literacy."

<sup>908</sup> Alliance for Childhood, Tech Tonic: Towards a New Literacy of Technology (College Park, Maryland: Alliance for Childhood, 2004).

<sup>909</sup> Asaravala, "Testing Your Tech Smarts."

## **CHAPTER 4**

### **RUBRICS AND INFORMATION LITERACY ASSESSMENT**

This chapter offers definitions of the term “rubric” and describes the major types of rubrics. It includes the benefits and limitations of using rubrics for the assessment of learning outcomes and provides examples of rubrics that focus on information literacy skills. Finally this chapter reviews guidelines for developing rubrics and tips for documenting results of rubric assessments.

#### **Definitions of Rubrics**

Rubrics are “descriptive scoring schemes” created by educators to guide analysis of student work.<sup>910</sup> Haffner writes, “In the educational literature and among the teaching and learning practitioners, the word ‘rubric’ is understood generally to connote a simple assessment tool that describes levels of performance on a particular task and is used to assess outcomes in a variety of performance-based contexts.”<sup>911</sup> Wiggins defines rubrics as:

A set of scoring guidelines for evaluating students’ work. Rubrics answer the following questions: By what criteria should performance be judged? Where should we look and what should we look for to judge performance success? What does the range in the quality of performance look like? How do we determine validity, reliability, and fairly what score should be given and what that score means? How should the different levels of quality be described and distinguished from one another?<sup>912</sup>

---

<sup>910</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>911</sup> Hafner, "Quantitative Analysis of the Rubric." 1509.

<sup>912</sup> Wiggins, Educative Assessment. 154.

Rubrics are often employed when educators must judge the quality of performances or constructed-response items<sup>913</sup> and they can be used across a broad range of subjects.<sup>914</sup> It is important to point out that “where and when a scoring rubric is used does not depend on the grade level, or subject, but rather on the purpose of assessment.”<sup>915</sup> Moskal clarifies that both pre-college and college instructors use rubrics;<sup>916</sup> however, not all college instructors use or even understand rubrics as a tool for assessing student learning. Hafner confirms:

It appears that the word “rubric” has little (if any) pedagogical relevance to the overwhelming majority of college-level and university-level teachers because of their academic appointment outside the education department and, in many cases, minimal preparation in teacher education; as a consequence, most professors in academia are usually unfamiliar with the popular pedagogical trends in “alternative” and “authentic” assessment of the past decade.<sup>917</sup>

Despite the fact that many college instructors are unfamiliar with rubrics, assessment experts proclaim rubrics “one of the most basic tools in the performance assessor’s kit.”<sup>918</sup>

Because institutions of higher education are increasingly asked to demonstrate student learning based on outcomes, rubrics merit increased attention. In fact, rubrics are a mainstay of outcomes-based education. Farmer

---

<sup>913</sup> Popham, Test Better, Teach Better. 95.

<sup>914</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>915</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>916</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>917</sup> Hafner, "Quantitative Analysis of the Rubric." 1509-1510.

<sup>918</sup> Wiggins, Educative Assessment. 153.



notes that, "Outcomes-based education typically includes rubrics that describe...performance at different levels of competencies."<sup>919</sup> Since outcomes-based education is likely to include performance assessments, rubrics are necessary to capture student learning. Marzano, Pickering, and McTighe point out, "As with most real-world tasks, performance tasks do not have a single correct answer; there are a variety of ways to successfully complete them. Consequently, students' performance of the tasks cannot be 'machine-scored,' but must be judged by one or more persons guided by well-defined criteria."<sup>920</sup> Herman, Aschbacher, and Winters describe importance of criteria within a rubric when assessing an assessment task:

Alternative assessments invite a wide...range of possible responses. Instead of judging responses as right or wrong, alternative assessments judge the quality of, and sometimes the process of, arriving at a complex response. To make such judgments and to ensure their validity, consistence, and fairness, we need criteria or scoring guidelines. Scoring criteria must be well-conceived, explicitly defined, and consistently applied. Well-specified criteria help to ensure that everyone understands what is expected.<sup>921</sup>

Because higher education needs to demonstrate that students are learning and because student learning is increasingly organized around outcomes, college instructors need to learn more about the use of rubrics to produce consistent and accurate assessments of student learning.

---

<sup>919</sup> Farmer, "Authentic Assessment of Information Literacy."

<sup>920</sup> Marzano, Pickering and McTighe, Assessing Student Outcomes. 29.

<sup>921</sup> John Herman, Pamela R. Aschbacher and Lynn Winters, "Setting Criteria," A Handbook for Student Performance in an Era of Restructuring, eds. Robert E. Blum and Judith A. Arter (Alexandria, Virginia: Association for Supervision and Curriculum Development, 1996). VI-4:3.

## Components of a Full Model Rubric

Rubric creators have a variety of rubric models from which to choose.<sup>922</sup> Indeed, four types of rubrics exist. They are checklists, advanced checklists, simple model, and full model.<sup>923</sup> A full model rubric is the most descriptive type of rubric and it holds the most promise as assessment tool capable of delivering detailed data about student learning. According to Bresciani, Zelna, and Anderson:

The full model contains the fullest descriptions of the list of criteria and more complete descriptions of the levels of mastery. Here, the intent is to get as detailed as possible so that the rubric can be applied by students to student, as well as applied by cocurricular specialists in other units. In addition, with this level of detail, the cocurricular specialist can even expect the student to self-evaluate their work.<sup>924</sup>

Rubrics of this type are usually formatted on a grid or table where the target indicators are listed down the left hand side and the levels of performance are listed across the top of the grid.<sup>925</sup>

The first of two components that comprise a full model rubric is criteria. In fact, Bresciani, Zelna, and Anderson note, “that the word ‘rubric’ cannot be defined without referencing criteria. Simply put, rubrics are an expansion of criteria.”<sup>926</sup> Popham also emphasizes the importance of criteria in a rubric. He writes, “The most important component in any rubric is the rubric’s evaluative

---

<sup>922</sup> Cyndi Phillip, "Clear Expectations: Rubrics and Scoring Guides," Knowledge Quest 31.2 (2002). 27.

<sup>923</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.

<sup>924</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 32.

<sup>925</sup> Daniel Callison, "Rubrics," School Library Media Activities Monthly 17.2 (2000). 34.

<sup>926</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 30.

criteria, which are the factors a scorer considers when determining the quality of a students' response.... Evaluative criteria are truly the guts of any rubric, because they lay out what it is that distinguishes students' winning response from their weak ones."<sup>927</sup> Wiggins defines criteria in this context as "the conditions a performance must meet to be successful."<sup>928</sup> Other authors define criteria as "the set of indicators, markers, guides, or a list of measures or qualities that will help [a scorer] know when a student has met an outcome."<sup>929</sup> Rubric criteria make the difference between asking students for meaningful and meaningless performances. Wiggins states that criteria should be "authentic, with points awarded or taken off for essential successes and errors, not for what is easy to count or observe."<sup>930</sup> Rubric criteria give educators the keys to recognize student mastery. According to Arter, criteria tell teachers what to look for in student performance "to determine progress, know when students are ready to move on to the next topic, or determine when mastery has occurred."<sup>931</sup> Arter states, "Good performance criteria are more than just a tool that teachers...can use to monitor students' ability to write or solve problems or be critical thinkers. They also help teachers and students conceptualize and define standards, especially in hard-to-define areas such as critical thinking [and] problem solving."<sup>932</sup> Criteria

---

<sup>927</sup> Popham, Test Better, Teach Better. 96.

<sup>928</sup> Wiggins, "Creating Tests Worth Taking." VI-5:1.

<sup>929</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 29.

<sup>930</sup> Wiggins, "Creating Tests Worth Taking." V-6:3.

<sup>931</sup> Arter, "Using Assessment." IV-10:4.

<sup>932</sup> Arter, "Using Assessment." IV-10:4.

based on standards have many benefits. Wiggins suggests that “because descriptors contain criteria and also often refer to standards, a good rubric makes possible valid and reliable—that is, criterion-referenced—discrimination of performances.”<sup>933</sup>

“Target indicators”, “quality definitions”, and “performance descriptors” are all phrases that describe the second component of rubrics. By whatever name, this part of a rubric “spell[s] out what is needed, with respect to each evaluative criterion, for a student’s response to receive a high rating versus a low rating on that criterion.”<sup>934</sup> By adding performance descriptors to a rubric, educators can go beyond just determining whether the criteria exists or not, and describe the levels of achievement in more detail.<sup>935</sup> The process of developing performance descriptors is a significant one. Arter explains, “If we truly want to ensure that all students attain the complex skills and thinking processes that will enable them to be successful and productive citizens, we must ensure that we know what success in these complex targets looks like and what constitutes satisfactory progress toward the ultimate target.”<sup>936</sup> Performance descriptors embody the achievement of Arter’s goal. She writes:

Performance criteria define and illustrate what to look for in student products or behaviors. They address concerns such as how to tell a good response from a poor one; how to be consistent in judging student performance across assignments, students, and time; how to define

---

<sup>933</sup> Wiggins, Educative Assessment. 154.

<sup>934</sup> Popham, Test Better, Teach Better. 96.

<sup>935</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 30.

<sup>936</sup> Arter, "Using Assessment." IV-10:6.

success; and how to capture student performance in a way to report this to parents.<sup>937</sup>

The best performance descriptors are written in detailed language. In a well-written rubric, “instead of stating that a student demonstrated the learning at a level of excellent or poor, excellent is described in detail for what the evaluator was looking to find. Poor is defined as well, as would any other middle achievement levels of the outcome that the evaluator would expect to see.”<sup>938</sup>

Callison notes the importance of the accessibility of the language used in a rubric. He recommends that “language is used, in terms that both the student and teacher understand, so that precise actions are defined for what the student must do to demonstrate a skill or proficiency at a certain level.”<sup>939</sup>

### **General and Task-Specific Rubrics**

Rubrics can be written to assess a specific task or a broader group of tasks.<sup>940</sup> Task-specific rubrics are designed to evaluate student performance on a single event.<sup>941</sup> In contrast, a general rubric can be used across similar performances.<sup>942</sup> Wiggins states that, “The more task-specific the rubric, the more valid the result; the more clear and simple the rubric, the greater the

---

<sup>937</sup> Arter, "Using Assessment." VI-2:1.

<sup>938</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 30.

<sup>939</sup> Callison, "Rubrics." 34.

<sup>940</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>941</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>942</sup> Judith Arter and Jay McTighe, Scoring Rubrics in the Classroom: Using Performance Criteria for Assessing and Improving Student Performance (Thousand Oaks, California: Corwin Press, 2000). 24.

reliability.”<sup>943</sup> Wiggins also weighs the benefits of task-specific rubrics and general rubrics:

Should a rubric be task-specific or more general? ... Somewhere in between really. We face a dilemma. Reliability is served by having a rubric unique to each task. But we have a feasibility and maybe reliability problem: it takes too much time and energy to design a rubric for every task. Furthermore, a specially designed rubric for each task may not be necessary.<sup>944</sup>

Moskal suggests that assessors may not have to choose one type over the other.

She writes, “Scoring rubrics may be designed to contain both general and task specific components.”<sup>945</sup> Wiggins agrees that, “The best rubrics are those that are sufficiently generic to relate to general goals beyond an individual task, but specific enough to enable useful and sound inferences about the task.”<sup>946</sup>

Ideally, these two types of rubrics merge, forming a rubric that assesses more than one performance without sacrificing validity or reliability.

### **Holistic and Analytic Rubrics**

Rubrics can also be described as holistic or analytic. While a rubric may combine task-specific and general attributes, a rubric cannot be holistic and analytic at the same time. Educators must choose which type to use for a given assessment task.

---

<sup>943</sup> Wiggins, "Creating Tests Worth Taking." VI-5:13.

<sup>944</sup> Wiggins, "Creating Tests Worth Taking." VI-5:9.

<sup>945</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>946</sup> Wiggins, Educative Assessment. 184.

A holistic rubric “requires the teacher to score the overall process or product as a whole, without judging the component parts separately.”<sup>947</sup> It gives one score for a whole product or performance based on an overall impression.<sup>948</sup> Holistic rubrics provide an “overall, single judgment of quality.”<sup>949</sup> Popham writes, “A holistic scoring strategy signifies that the scorer must attend to how well a student’s response satisfies all the evaluative criteria in the interest of forming a general, overall evaluation of the response based on all criteria considered in concert.”<sup>950</sup>

Holistic rubrics have both advantages and disadvantages. For large-scale projects, holistic rubrics are often preferred because they are faster, and therefore, less expensive to use.<sup>951</sup> Arter and McTighe recommend holistic rubrics for simple products or performances, particularly ones with only one important criterion to assess.<sup>952</sup> Moskal points out that holistic rubrics are sometimes preferred when the criteria that mark a product or performance are difficult to distinguish or when they overlap.<sup>953</sup> Arter and McTighe also note that holistic rubrics are useful for “getting a quick snapshot of overall quality or

---

<sup>947</sup> Anthony J. Nitko, Educational Assessment of Students (Englewood Cliffs, New Jersey: Prentice Hall, 1996). 226.

<sup>948</sup> Arter and McTighe, Scoring Rubrics. 18.

<sup>949</sup> Arter and McTighe, Scoring Rubrics. 18.

<sup>950</sup> Popham, Test Better, Teach Better. 96.

<sup>951</sup> Popham, Test Better, Teach Better. 97.

<sup>952</sup> Arter and McTighe, Scoring Rubrics. 21.

<sup>953</sup> Moskal, "Scoring Rubrics: What, When, and How?."

achievement.”<sup>954</sup> This is especially important for large-scale assessments where it is necessary to assess large numbers of student responses.<sup>955</sup> Mertler states that holistic rubrics are often used when the focus is on a summative assessment, since they provide only limited feedback.<sup>956</sup> This points to one disadvantage of holistic rubrics. They 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.”<sup>957</sup> Rockman’s information literacy rubric<sup>958</sup> is an example of a holistic rubric that has these strengths and weaknesses (see Figure 4.6).

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.”<sup>959</sup> Analytic rubrics allow for separate evaluations of each factor along a different descriptive scale.<sup>960</sup> Then, assessors sum the individual scores to form a total

---

<sup>954</sup> Arter and McTighe, Scoring Rubrics. 21.

<sup>955</sup> Arter and McTighe, Scoring Rubrics. 21.

<sup>956</sup> Craig A. Mertler, "Designing Scoring Rubrics for Your Classroom," Practical Assessment, Research and Evaluation 7.25 (2001).

<sup>957</sup> Arter and McTighe, Scoring Rubrics. 21.

<sup>958</sup> Ilene F. Rockman, Rubrics for Assessing Information Competence in the California State University, 2002, April 10 2005 <[http://www.calstate.edu/LS/1\\_rubric.doc](http://www.calstate.edu/LS/1_rubric.doc)>.

<sup>959</sup> Arter and McTighe, Scoring Rubrics. 18.

<sup>960</sup> Moskal, "Scoring Rubrics: What, When, and How?."



score.<sup>961</sup> Popham writes, “An analytic approach to scoring requires a scorer to make a criterion-by-criterion judgment for each of the evaluative criteria, and then amalgamate those per-criterion ratings into a final score (this is often done via a set of predetermined, possibly numerical, rules).”<sup>962</sup> Because of this part-to-whole approach, analytic rubrics offer a multidimensional assessment approach.<sup>963</sup> It is interesting to note that there is no agreed upon number of traits to include in an analytic rubric. The number of dimensions an analytic rubric can score depends on the complexity of the skill or product.<sup>964</sup>

Like holistic rubrics, there are both advantages and disadvantages to using an analytic rubric approach to assessment. According to Mertler, analytic rubrics are preferred when a focused response to stakeholders is required.<sup>965</sup> Analytic rubrics provide an advantage over holistic rubrics when the product or performance may result in “one or two acceptable responses.”<sup>966</sup> They also are better suited for “judging complex performances (e.g. research process) involving several significant dimensions.... By breaking such performances down into traits, raters (including students and teachers) can more readily grasp the essential components of quality.”<sup>967</sup> Mertler states that the advantages of

---

<sup>961</sup> Nitko, Educational Assessment of Students. 266.

<sup>962</sup> Popham, Test Better, Teach Better. 96.

<sup>963</sup> Mertler, "Designing Scoring Rubrics."

<sup>964</sup> Arter and McTighe, Scoring Rubrics. 21.

<sup>965</sup> Mertler, "Designing Scoring Rubrics."

<sup>966</sup> Mertler, "Designing Scoring Rubrics."

<sup>967</sup> Arter and McTighe, Scoring Rubrics. 22.

analytic rubrics are “substantial,” also citing the feedback they offer to students that simply “does not happen” when holistic rubrics are used.<sup>968</sup> He notes that the feedback is detailed enough to create a “profile” of student strengths and weaknesses.<sup>969</sup> Arter and McTighe suggest that such rich feedback can be used to target future instruction to specific areas of need.<sup>970</sup> They write, “From an instructional perspective, analytical trait rubrics help students come to better understand the nature of quality work since they identify the important dimensions of a product or performance.”<sup>971</sup> Popham agrees that analytic rubrics are more diagnostic and notes that they allow an assessment of student performance per-criterion.<sup>972</sup> However, all of these advantages have a cost. Analytical rubrics take more time to create and use. Arter and McTighe write, “After all, you have more to discern.”<sup>973</sup> Mertler also notes the problem of time. He writes:

The use of analytical rubrics can cause the scoring process to be substantially slower, mainly because assessing several different skills or characteristics individually requires a teacher to examine the product several times. Both their construction and use can be quite time-consuming. A general rule of thumb is that an individual’s work should be examined a separate time for each of the specific performance tasks or scoring criteria.<sup>974</sup>

---

<sup>968</sup> Mertler, “Designing Scoring Rubrics.”

<sup>969</sup> Mertler, “Designing Scoring Rubrics.”

<sup>970</sup> Arter and McTighe, Scoring Rubrics. 22.

<sup>971</sup> Arter and McTighe, Scoring Rubrics. 22.

<sup>972</sup> Popham, Test Better, Teach Better. 97.

<sup>973</sup> Arter and McTighe, Scoring Rubrics. 23.

<sup>974</sup> Mertler, “Designing Scoring Rubrics.”

Finally, Arter and McTighe acknowledge that analytic rubrics can result in lower agreement among raters, at least initially, because of the greater number of criteria that are examined.<sup>975</sup>

The decision to use a holistic or analytic rubric is one that ultimately depends on the product or performance to be assessed, the criteria to be observed,<sup>976</sup> and the purpose of the assessment. Arter and McTighe state that, "Any good rubric, holistic or analytic...will cover all the essential features of a performance...the only question is whether to leave the whole ball of wax as a ball (holistic scoring) or to group similar features together and slice the ball up into traits."<sup>977</sup> According to Mertler, the most important consideration 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."<sup>978</sup> Educators might also consider the time requirements of a particular assessment scenario.<sup>979</sup>

### **Benefits of Rubric Assessment**

The benefits of rubric assessment are numerous.<sup>980</sup> Rubrics have

---

<sup>975</sup> Arter and McTighe, Scoring Rubrics. 23.

<sup>976</sup> Mertler, "Designing Scoring Rubrics."

<sup>977</sup> Arter and McTighe, Scoring Rubrics. 20.

<sup>978</sup> Mertler, "Designing Scoring Rubrics."

<sup>979</sup> Mertler, "Designing Scoring Rubrics."

<sup>980</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 30.

instructional advantages over other assessment approaches, including the direct benefits to students, the benefits of stating agreed upon values, and the benefits of detailed result data. Rubrics also offer advantages to assessors focusing on standards-based education, student learning across multiple programs, and cost.

Many educational researchers have demonstrated the instructional benefits of scoring rubrics.<sup>981</sup> They find that rubrics provide a “map and guide for student assessment.”<sup>982</sup> Popham calls rubrics “instructional illuminators” and emphasizes that “appropriately designed rubrics can make an enormous contribution to instructional quality.”<sup>983</sup> Popham continues, “A properly fashioned rubric can help teachers teach much more effectively and help students learn much more effectively, too.”<sup>984</sup> According to Lichtenstein, rubrics make learning easier and more effective because, “people usually gain more from a lesson when they are told what it is they are supposed to learn.”<sup>985</sup> Phillip notes the match between rubric assessment and learning theory. She writes, “Learning theory supports the idea that we retain the most when we are actively involving all of our senses in a doing mode. By creating, sharing, and accomplishing the criteria set by a rubric, the student is in charge of his or her own learning and

---

<sup>981</sup> Robin Tierney and Marielle Simon, "What's Still Wrong with Rubrics: Focusing on the Consistency of Performance Criteria across Scale Levels," Practical Assessment, Research, and Evaluation 9.2 (2004).

<sup>982</sup> Callison, "Rubrics." 36.

<sup>983</sup> W. James Popham, "What's Wrong--and What's Right--with Rubrics," Educational Leadership 55.2 (1997). 75.

<sup>984</sup> Popham, Test Better, Teach Better. 95.

<sup>985</sup> Lichtenstein, "Informed Instruction." 28.

assessment.”<sup>986</sup> Pausch and Popp suggest that rubrics are more valuable for learners than other assessment tools because they emphasize “understanding rather than memorization, ‘deep’ learning rather than ‘surface’ learning.”<sup>987</sup> According to Pausch and Popp, “Assessing in ways that foster ‘deep’ learning is important because research shows students learn what they expect will be assessed.”<sup>988</sup> Popham agrees that rubrics guide both teachers and students to deep mastery of skills.<sup>989</sup> Resnick concurs and issues a general call for attention to rubric assessment. She writes, “As educators, it is our job to help students understand how to construct their own learning and thereby continue to be life-long learners. The use of rubrics as a teaching and learning tool can play an integral part in attaining this goal.”<sup>990</sup>

While researchers agree that rubrics are generally useful in teaching and learning processes, there are specific benefits of rubric usage. Most importantly, rubrics benefit students in several ways. Rubrics allow students to understand the expectations of their instructors. Rubrics provide direct feedback to students about what they have learned and what they have yet to learn. They also can facilitate student self-evaluation. Rubrics are also used to make rankings, ratings, and grades more meaningful by revealing what educators expect

---

<sup>986</sup> Phillip, "Clear Expectations: Rubrics and Scoring Guides." 26.

<sup>987</sup> Pausch and Popp, "Assessment of Information Literacy."

<sup>988</sup> Pausch and Popp, "Assessment of Information Literacy."

<sup>989</sup> Popham, Test Better, Teach Better. 101.

<sup>990</sup> Resnick, Education and Learning to Think.

students to know or do.<sup>991</sup> Phillip explains, "The value of producing rubrics and presenting them to students, or preparing them with the students, lies in the fact that expectations are clearly stated. The student will know what is required for the fulfillment of written, oral, or visual projects."<sup>992</sup> She notes rubrics should provide students with the quality criteria, guidelines, and standards of evaluation connected to an assignment.<sup>993</sup> Putting such criteria into a rubric helps both students and teachers understand why a particular grade is given to a student's product or performance.<sup>994</sup> It helps students understand "what they have 'missed' and what they have yet to learn."<sup>995</sup> Callison affirms that organizing goals for student performance on a grid or map gives students "a more clear visual of their progress and what is necessary to achieve a higher rating than is reflected in simple letter grades."<sup>996</sup> Hafner agrees that rubrics provide quick and clear summaries of performance levels and emphasizes the importance of the top level of a rubric. He writes, "Importantly, the top level of the rubric communicates what exemplary work should look like and, as such, involves the student in constructive learning and self-evaluation."<sup>997</sup>

---

<sup>991</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.

<sup>992</sup> Phillip, "Clear Expectations: Rubrics and Scoring Guides." 26.

<sup>993</sup> Phillip, "Clear Expectations: Rubrics and Scoring Guides." 26.

<sup>994</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.

<sup>995</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.

<sup>996</sup> Callison, "Rubrics." 35.

<sup>997</sup> Hafner, "Quantitative Analysis of the Rubric." 1510.

Providing criteria and quality standards to students allows them to understand their assignments and how they are graded. Lazzaro describes the ways in which this understanding can empower students. He notes that students must understand the content of scoring rubrics so they can deliver their best work.<sup>998</sup> Herman, Aschbacher, and Winters describe the importance of discussing scoring rubrics with students:

Public discussions help students to internalize the standards and 'rules' they need to become independent learners.... Examples of what constitutes good work engage students in the work itself and in judgments about their work. Public discussions of quality and criteria inform students during the formative period of instruction, not simply at the end of a unit or course when it is too late to make improvements. Furthermore, discussions of criteria also help students see the perspectives of their teachers, their peers, and sometimes even of the experts in the field.<sup>999</sup>

Shepard asserts that providing students with access to evaluation criteria is required for "basic fairness".<sup>1000</sup> She believes that students should know the rules for how their products and performances will be judged. She adds, "Giving students the opportunity to get good at what it is that the standards require speaks to a different and even more fundamental sense of fairness."<sup>1001</sup>

Shepard also cites Frederikson and Collin's use of the term "transparency" to describe the need for students to understand how their work will be scored.<sup>1002</sup>

---

<sup>998</sup> William L. Lazzaro, "Empowering Students with Instructional Rubrics," A Handbook for Student Performance in an Era of Restructuring, eds. Robert E. Blum and Judith A. Arter (Alexandria, Virginia: Association for Supervision and Curriculum Development, 1996). VI-3:3.

<sup>999</sup> Herman, Aschbacher and Winters, "Setting Criteria." VI-4:3.

<sup>1000</sup> Shepard, "The Role of Assessment in a Learning Culture." 11-12.

<sup>1001</sup> Shepard, "The Role of Assessment in a Learning Culture." 11-12.

<sup>1002</sup> Shepard, "The Role of Assessment in a Learning Culture." 11.

Shepard believes that students should understand scoring criteria so well that they can evaluate their own work exactly as their teachers would.<sup>1003</sup> She further cites Frederikson and Collin in stating that rubrics should make students metacognitively aware of the characteristics of good work.<sup>1004</sup> By truly understanding the criteria with which their work will be assessed, students should “perform better than learners who do not have this knowledge.”<sup>1005</sup>

Rubrics that are well constructed and shared with students not only allow students to understand what is asked of them up front, they also have the added benefit of providing timely<sup>1006</sup> and detailed feedback to students.<sup>1007</sup> Bresciani, Zelna and Anderson state that rubrics can be used as teaching tools.<sup>1008</sup> When used as tools for feedback, “students begin to understand what it is they are or are not learning, and are or are not able to demonstrate what they know and can do. When students begin to see what they are not learning, they can take more responsibility for their learning.”<sup>1009</sup> Similarly, rubrics can be used to provide feedback to students about their improvement.<sup>1010</sup> Callison also points out that

---

<sup>1003</sup> Shepard, "The Role of Assessment in a Learning Culture." 11.

<sup>1004</sup> Shepard, "The Role of Assessment in a Learning Culture." 11.

<sup>1005</sup> Lazzaro, "Empowering Students with Instructional Rubrics." VI-3:3.

<sup>1006</sup> Dannelle D. Stevens and Antonia Levi, Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback, and Promote Student Learning (Sterling, Virginia: Stylus, 2005). 17.

<sup>1007</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1008</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.

<sup>1009</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.

<sup>1010</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.



rubrics make feedback easy to convey and that feedback can “provide a benchmark from one lesson to the next.”<sup>1011</sup>

As students become comfortable with rubric assessment, they can use rubrics independently for self-evaluation.<sup>1012</sup> Students can evaluate themselves and their peers with rubrics, and this process helps “students identify their own learning and development or absence thereof.”<sup>1013</sup> Bresciani, Zelna and Anderson summarize, “If you can explain in detail to students what they are expected to get out of a particular experience or series of activities, the students begin to see value in the activity and can then assist in their own evaluations as well as assume greater responsibility for their own performances.”<sup>1014</sup> When students participate in self-assessment, there are many benefits. Stevens and Levi state that rubrics “encourage students to think critically about their own learning.”<sup>1015</sup> According to Shepard, students benefit from increased responsibility for learning and a more collaborative relationship between teachers and students.<sup>1016</sup> Shepard cites Gipps, Klenowski, and Wiggins in her argument that self-evaluation allows students to gain ownership, be honest about their

---

<sup>1011</sup> Callison, "Rubrics." 35.

<sup>1012</sup> Rosemarie Bernier, "Making Yourself Indispensable by Helping Teachers Create Rubrics," CSLA Journal 27.2 (2004). 25.

<sup>1013</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.

<sup>1014</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.

<sup>1015</sup> Stevens and Levi, Introduction to Rubrics. 21.

<sup>1016</sup> Shepard, "The Role of Assessment in a Learning Culture." 12.

work, become more interested in learning than grades, and be held to higher standards, all of which result in increased investment in learning.<sup>1017</sup>

Not only can students assess themselves using rubrics, they can also work cooperatively with peers to assess each other. One question that might arise from this practice is, “How valid and reliable are rubrics in the hands of students?” Hafner conducted a study to investigate the use of rubrics by students whom he considered “pedagogically naïve raters.”<sup>1018</sup> Using the instructor’s scores as a comparison to measure validity, he determined that student scores were able to predict the instructor’s scores accurately.<sup>1019</sup> Hafner also found “significant inter-rater reliability” across the study period of 3 years.<sup>1020</sup> Furthermore, he discovered that the ability to use the rubric was “gender neutral” and had no bearing on students’ overall academic strength in the course.<sup>1021</sup> Hafner concluded that using rubrics for peer assessment “provides an effective teaching and learning strategy.”<sup>1022</sup>

Rubric assessment provides another important instructional benefit—the opportunity to discuss and determine agreed upon values of student learning. Callison writes, “Rubrics are texts that are visible signs of agreed-upon values. They cannot contain all the nuances of the evaluation community’s values, but

---

<sup>1017</sup> Shepard, “The Role of Assessment in a Learning Culture.” 12.

<sup>1018</sup> Hafner, “Quantitative Analysis of the Rubric.” 1510.

<sup>1019</sup> Hafner, “Quantitative Analysis of the Rubric.” 1514, 1520.

<sup>1020</sup> Hafner, “Quantitative Analysis of the Rubric.” 1518-1519.

<sup>1021</sup> Hafner, “Quantitative Analysis of the Rubric.” 1519.

<sup>1022</sup> Hafner, “Quantitative Analysis of the Rubric.” 1526.

they do contain the central expressions of those values.”<sup>1023</sup> Stevens and Levi list the facilitation of communication with others as a key reason to use rubrics.<sup>1024</sup> Bresciani, Zelna and Anderson confirm that rubrics “make public key criteria that students can use in developing, revising, and judging their own work.”<sup>1025</sup> They also point out that once rubrics are developed, they can be used to norm educators’ expectations and to bring them in line with the original vision for student learning.<sup>1026</sup> They can also be used to make students full participants in the assessment process. Rubrics allow the possibility of including students in the development of standards for a product or performance.<sup>1027</sup> Bernier points out that “helping teachers create rubrics helps students meet those content standards.”

Using rubrics based on agreed upon values has practical advantages for consistent scoring. First, “if applied correctly, rubrics come close to assuring that inadequate, satisfactory, and excellent mean the same thing on the same skill set from one group of students to a similar group regardless of who makes the evaluation.”<sup>1028</sup> It also is a benefit for program development. Bresciani, Zelna and Anderson write:

---

<sup>1023</sup> Callison, "Rubrics." 36.

<sup>1024</sup> Stevens and Levi, Introduction to Rubrics. 23.

<sup>1025</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 30.

<sup>1026</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.

<sup>1027</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.

<sup>1028</sup> Callison, "Rubrics." 35.

If faculty and cocurricular specialists join together to precisely articulate criteria and levels of performance for student learning and development, they can then work more diligently on improving the programs that would provide such learning and development experiences, because the faculty and cocurricular specialists share the desired learning and development outcome.<sup>1029</sup>

Finally, the values that educators, assessors, and even students agree upon can be shared with parents and used to influence administrators.<sup>1030</sup>

Rubric assessment offers a third instructional benefit—data full of rich description. Rubrics provide “evaluators and those whose work is being evaluated with rich and detailed descriptions of what is being learned and what is not.”<sup>1031</sup> This descriptive data can be used to document how educators or program administrators improve instruction.<sup>1032</sup> Furthermore, the data that results from rubric assessment is so detailed and well-defined that it “combats accusations that evaluators do not know what they are looking for in learning and development.”<sup>1033</sup> The level of detail found in rubrics helps prevent inaccuracy of scoring<sup>1034</sup> and bias.<sup>1035</sup> Rubrics clarify schemes for assessment ahead of time, and therefore reduce subjectivity in grading.<sup>1036</sup> Since rubrics guide teachers to

---

<sup>1029</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.

<sup>1030</sup> Bernier, "Making Yourself Indispensable." 25.

<sup>1031</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 30.

<sup>1032</sup> Bernier, "Making Yourself Indispensable." 25.

<sup>1033</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 30.

<sup>1034</sup> Popham, Test Better, Teach Better. 95.

<sup>1035</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.

<sup>1036</sup> Moskal, "Scoring Rubrics: What, When, and How?."

focus on essential criteria,<sup>1037</sup> they can grade student products and performances more easily and objectively.<sup>1038</sup> Callison indicates the importance of this benefit for librarians, stating that rubric assessment “is more likely to be reasonably objective and consistent from lesson to lesson and from student to student, especially useful in team teaching situations that involve collaboration among library media specialists and other teachers.”<sup>1039</sup> Because rubrics are easy to use and easy to explain, they also generate data that is easy to understand, defend, and convey.<sup>1040</sup>

Additional benefits of a rubric approach to assessment include a focus on standards, transferability across time or multiple programs, and cost. According to Bresciani, Zelna and Anderson, rubrics teach students “the standards of the discipline or the standards of the cocurricular learning and development experience.”<sup>1041</sup> They help determine the extent to which students achieve standards.<sup>1042</sup> Rubrics can measure achievement of standards from a formative perspective or an “ecological” one, revealing how students “apply knowledge in authentic situations.”<sup>1043</sup> This is significant because institutional accreditors often

---

<sup>1037</sup> Callison, "Rubrics." 35.

<sup>1038</sup> Bernier, "Making Yourself Indispensable." 25.

<sup>1039</sup> Callison, "Rubrics." 35.

<sup>1040</sup> Heidi Goodrich Andrade, "Using Rubrics to Promote Thinking and Learning," Educational Leadership 57.5 (2000). 14.

<sup>1041</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.

<sup>1042</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1043</sup> Pausch and Popp, "Assessment of Information Literacy."

seek “multiple dimensions of student performance,” not just a summative grade.<sup>1044</sup>

Rubrics also can be used to assess student learning across time and programs. Rubrics are flexible and can be used to assess students, activities, or programs.<sup>1045</sup> Educators can use the same or similar rubrics over the course of time, even an academic career, to assess student progress toward learning goals.<sup>1046</sup> Rubrics can also be used to assess individual students, all students in a program, or students across multiple programs.<sup>1047</sup> This is important for educators and program administrators who need to capture the learning of a student population that flows from program to program.<sup>1048</sup> Bresciani, Zelna and Anderson provide advice on how to conduct a cross-program assessment plan:

We may not be able to have these students long enough to assess what they have learned from us, or we may only capture a point in their learning and development and then they move onto another cocurricular specialist's program. Thus, using similar rubrics across programs for problem solving, collaboration, reflection, and other such shared outcomes may be possible, if all can agree on some basic criteria and levels of performance.... When using the same rubric across programs, units or division lines, it is important to have some basic conformity. Following are some basic agreements that are advantageous to have in place when sharing rubrics across organizational lines: Agree on the outcome you are assessing with the rubric. Agree on a method of data collection such as an essay or presentation. Agree on the meaning for the outcome and definition. In other words, agree on how you know the outcome is met and

---

<sup>1044</sup> Council for Higher Education Accreditation, "Student Learning Outcomes Workshop." 2.

<sup>1045</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.

<sup>1046</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.

<sup>1047</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.

<sup>1048</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 35.

what it will look like when you see it met. Agree on the systematic implementation of the assignments, activities, projects, and the rubric.<sup>1049</sup>

Two more benefits of rubric assessment are low cost and demonstrated effectiveness. Although rubrics require an investment in time for conversation and norming, they are inexpensive to design and implement.<sup>1050</sup> Rubrics also have been used extensively in the assessment of constructed-response test items and overall writing instruction and they have been found to “go along way toward neutralizing the drawbacks of constructed-response assessments while maximizing the advantages”<sup>1051</sup> and “contribute wonderfully to accurate scoring and successful [writing] instruction.”<sup>1052</sup> Indeed Stevens and Levi contend, “The main reason we don’t use rubrics more often is simply because most of us have been unaware of them. Rubrics were not part of our own experience as students, and most of us find that we often teach as we were taught.”<sup>1053</sup>

### **Limitations of Rubric Assessment**

Like other assessment tools, there are limitations associated with rubric assessment. Many of the limitations of a rubric approach to assessment are rooted in poor rubric construction. Not all rubrics are well-written,<sup>1054</sup> and crafting

---

<sup>1049</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 35-36.

<sup>1050</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 37.

<sup>1051</sup> Popham, Test Better, Teach Better. 101.

<sup>1052</sup> Popham, Test Better, Teach Better. 97.

<sup>1053</sup> Stevens and Levi, Introduction to Rubrics. 17.

<sup>1054</sup> Popham, Test Better, Teach Better. 95.

a good rubric requires time, practice, and revision.<sup>1055</sup> Tierney and Simon caution that, unfortunately, “the most accessible rubrics, particularly those available on the Internet, contain design flaws that not only affect their instructional usefulness, but also the validity of their results.”<sup>1056</sup>

One frequent contributor to poor rubric construction is the difficulty of balancing between generalized wording, which increases the amount of different tasks to which the rubric can be applied, and detailed description, which provides a level of specificity that ensures greater reliability.<sup>1057</sup> Popham describes rubrics that are too detailed:

This type of rubric contains evaluative criteria that refer to the particular task that the student has been asked to perform, not the skill that task was created to represent.... The rubric’s evaluative criteria, focused exclusively on the particular task, would have no relevance to the evaluation of students’ responses to other task representing the cognitive skill being measured.<sup>1058</sup>

Sometimes, the problem of too much detail is compounded by too much length. Popham describes these rubrics as having “loads and loads of evaluative criteria, each with its own extremely detailed set of quality descriptions. These rubrics are just blinking long. Because few teachers and students have the patience or time to wade through rubrics of this type, these rubrics don’t help teachers teach

---

<sup>1055</sup> Callison, "Rubrics." 35.

<sup>1056</sup> Tierney and Simon, "What's Still Wrong with Rubrics."

<sup>1057</sup> Tierney and Simon, "What's Still Wrong with Rubrics."

<sup>1058</sup> Popham, Test Better, Teach Better. 98.



or students learn.”<sup>1059</sup> Popham notes that the best rubrics are concise. He writes:

Brevity is the best way to ensure that a rubric is read and used by both teachers and students. Ideally each evaluative criterion should have a brief, descriptive label. And...it usually makes sense for the teacher to conjure up a short, plain-talk student version of any decent skill-focused rubric.<sup>1060</sup>

Another contribution to poor rubric construction is vague wording. When rubric criteria or performance descriptions are too general, a rubric becomes useless.<sup>1061</sup> Popham writes, “Numerous rubrics have criteria so amorphous they are almost laughable.... In essence, these overly general criteria allow both teachers and students to conclude that really good student responses are, well, really good. And, of course, really bad student responses are—you guessed it—really bad.”<sup>1062</sup> Such rubrics provide little more meaning than an A through F letter grade, and therefore don't help teachers or students to improve.<sup>1063</sup>

Inconsistency is another characteristic of poorly constructed rubrics. Tierney and Simon note that many ready-made, available rubrics have consistency flaws in the treatment of criteria across performance levels.<sup>1064</sup> They state, “Unfortunately, many rubrics are still not instructionally useful because of inconsistencies in the descriptions of performance criteria across their scale

---

<sup>1059</sup> Popham, Test Better, Teach Better. 98.

<sup>1060</sup> Popham, Test Better, Teach Better. 99.

<sup>1061</sup> Popham, "What's Wrong--and What's Right--with Rubrics." 73.

<sup>1062</sup> Popham, "What's Wrong--and What's Right--with Rubrics." 73.

<sup>1063</sup> Popham, Test Better, Teach Better. 98.

<sup>1064</sup> Tierney and Simon, "What's Still Wrong with Rubrics."

levels.”<sup>1065</sup> They blame inconsistency on “the fact that consistency has not been discussed extensively in relation to rubric design.”<sup>1066</sup> They recommend using parallel language to combat this flaw.<sup>1067</sup> However, they warn that, without greater consistency, rubrics cannot “fulfill their educational ideal.”<sup>1068</sup>

Researchers identify a few additional problems that can result in poor rubric construction. Tierney and Simon point out that performance levels should be clearly differentiated and descriptors need to be clear, otherwise students and raters may misinterpret a rubric.<sup>1069</sup> Callison notes problems with rubrics that emphasize quantity, rather than quality.<sup>1070</sup> He suggests that rubrics should not judge student performance on how many times a student does something, but on how well he or she does it.<sup>1071</sup> Finally, some rubrics describe the lower levels of student performance in negative terms, resulting in a “dichotomous” tone.<sup>1072</sup>

Tierney and Simon write:

Students who find themselves on the lower part of the scoring rubric may not be motivated to progress with this type of feedback.... This doesn't mean that words, such as none, not or seldom, should always be avoided in rubric design, but that their use should represent one end of a continuous and consistent scale without undue negativity. However, when rubrics are not modified to reflect a positive continuum, they may

---

<sup>1065</sup> Tierney and Simon, "What's Still Wrong with Rubrics."

<sup>1066</sup> Tierney and Simon, "What's Still Wrong with Rubrics."

<sup>1067</sup> Tierney and Simon, "What's Still Wrong with Rubrics."

<sup>1068</sup> Tierney and Simon, "What's Still Wrong with Rubrics."

<sup>1069</sup> Tierney and Simon, "What's Still Wrong with Rubrics."

<sup>1070</sup> Callison, "Rubrics." 36.

<sup>1071</sup> Callison, "Rubrics." 36.

<sup>1072</sup> Tierney and Simon, "What's Still Wrong with Rubrics."

perpetuate low expectations for certain students rather than promote learning.<sup>1073</sup>

An area of difficulty surrounding rubric assessment that is not directly related to rubric content or construction is the issue of assigning grades. Mertler discourages educators from converting rubric scores into percentages, arguing instead for a logical, rather than mathematical, approach.<sup>1074</sup> He emphasizes that there is no “correct way” to convert rubric scores to grades, although he provides an example of one way to do so. Instead he suggests that educators “find a system of conversion that works for them and fits comfortably into their individual system of reporting student performance.”<sup>1075</sup>

The last limitation of rubric assessment is time. While creating rubrics is inexpensive monetarily, some teachers find the process time-consuming.<sup>1076</sup> Part of that perception might be due to lack of familiarity or expertise; teachers don’t always know how to make a rubric and so they believe the process will take too much time.<sup>1077</sup> Prus and Johnson acknowledge the potential cost of time required to create a rubric, but feel that the advantages outweigh the costs.<sup>1078</sup> They write: “As in virtually all other domains of human assessment, there is a consistently inverse correlation between the quality of measurement methods

---

<sup>1073</sup> Tierney and Simon, "What's Still Wrong with Rubrics."

<sup>1074</sup> Mertler, "Designing Scoring Rubrics."

<sup>1075</sup> Mertler, "Designing Scoring Rubrics."

<sup>1076</sup> Tierney and Simon, "What's Still Wrong with Rubrics."

<sup>1077</sup> Bernier, "Making Yourself Indispensable." 25.

<sup>1078</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 25.

and their expediency; the best methods usually take longer and cost more faculty time, student effort, and money.”<sup>1079</sup> Stevens and Levi argue that rubrics actually make grading easier and faster by “establishing performance anchors, providing detailed, formative feedback,...supporting individualized, flexible, formative feedback,...and conveying summative feedback.”<sup>1080</sup>

### **Rubric Development**

Developing a rubric to assess student learning requires a multi-step approach. Any kind of “assessment requires academic organizations...to: make expectations and standards for quality explicit and public; systematically gather evidence on how well performance matches those expectations and standards; analyze and interpret the evidence; and use the resulting information to document, explain, and improve performance.”<sup>1081</sup> However, the process of developing a rubric requires additional steps, beginning with initial preparation.

Rubric development works best if a team approach is adopted early. The Council for Higher Education Accreditation recommends forming a four to eight member team that includes “both functional area staff members and representatives from other campus constituencies” such as faculty, staff, and students.<sup>1082</sup> They recommend team training so that all members understand the

---

<sup>1079</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 25.

<sup>1080</sup> Stevens and Levi, Introduction to Rubrics. 73.

<sup>1081</sup> Smith, New Roles and Responsibilities for the University Library.

<sup>1082</sup> Council for Higher Education Accreditation, "Student Learning Outcomes Workshop."

task and the standards that will be applied to the task.<sup>1083</sup> This team also needs to accomplish other goals at the outset of the assessment project. First, team members need to acknowledge their goals and purposes in conducting assessment, determine a common language, acknowledge political realities, articulate expectations, and review what has already been accomplished in the target area.<sup>1084</sup> After accomplishing these preliminary goals, the team should act. Rather than preparing indefinitely, team members should recognize that assessment is an iterative process<sup>1085</sup> and that what doesn't go perfectly in the first cycle can be learned from and improved in the next. Carter points out that, "Outcomes assessment is an ongoing process. One need not wait for the perfect opportunity, the perfect instrument, or the perfect time. A modest study that is well designed and collects hard data can provide valuable information. One can learn from each experience and refine and improve assessment procedures with each effort."<sup>1086</sup>

After sufficient preparation has occurred, the next step in developing a rubric to assess student learning is to articulate a student learning outcome to assess.<sup>1087</sup> In some cases, outcomes may be borrowed from professional bodies or institutional documents. In others, outcomes have to be generated and composed through an organic process within a program. After an outcome is

---

<sup>1083</sup> Council for Higher Education Accreditation, "Student Learning Outcomes Workshop."

<sup>1084</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 9.

<sup>1085</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 9.

<sup>1086</sup> Carter, "'Doing the Best You Can with What You Have'." 41.

<sup>1087</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 33.

identified and articulated, the next step is to determine “what meeting the outcome looks like.”<sup>1088</sup> Farmer recommends that educators ask themselves how they know when an outcome has been met.<sup>1089</sup> She writes, “What does it look like? This is key, as you need to articulate exactly what you are looking for and how you will know it has been met.”<sup>1090</sup> Educators can also refer to faculty colleagues and scholarly literature for assistance envisioning what the outcome looks like when it has been achieved.<sup>1091</sup>

At this point, the assessment team should determine a method for collecting evidence that the outcome has been met.<sup>1092</sup> Because different methods will impact the way learning is demonstrated, this step should be accomplished before the fourth step is attempted.<sup>1093</sup> Once the method for capturing evidence of student learning is established, the team can use the descriptions of what learning looks like into a rubric.

The next major step of rubric development requires assessment team members to digest descriptions of student learning into the criteria that will form the basic structure of a rubric.<sup>1094</sup> Using scholarly literature, professional experience, and feedback from colleagues, the team should attempt to articulate

---

<sup>1088</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 33.

<sup>1089</sup> Farmer, "Authentic Assessment of Information Literacy."

<sup>1090</sup> Farmer, "Authentic Assessment of Information Literacy."

<sup>1091</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 34.

<sup>1092</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 34.

<sup>1093</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 34.

<sup>1094</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 34.

evaluation criteria. Bresciani, Zelna and Anderson note that this can be a particularly difficult task.<sup>1095</sup> They write:

Often, we witness that cocurricular specialists do not trust themselves to articulate what they have known for years and what they do know because they are the expert in identifying what they want students to be able to know and do in their particular programs. They are experts because they have read and applied the literature, have learned from their students, and have engaged in ongoing professional development. So, trust yourself, write it down. You will have plenty of time for it to be critiqued and improved by your cocurricular and academic colleagues.<sup>1096</sup>

After undergoing reflection and revision, these underlying characteristics of student learning need to be identified and stated in the final rubric.<sup>1097</sup> Each of these criteria should “represent a key attribute of the skill being assessed.”<sup>1098</sup> Furthermore, each criterion should apply to any skill reflective task and be teachable.<sup>1099</sup> Popham states that, “Teachers need to be able to teach students to master an important skill. Thus each evaluative criterion on a good skill-focused rubric should be included only if an affirmative answer is given to this question: ‘Can I get my students to be able to use this evaluative criterion in judging their own mastery of the skill I’m teaching them?’”<sup>1100</sup> Popham also recommends that rubrics include only a “handful” of criteria.<sup>1101</sup> He suggests that rubrics “incorporate only the most important evaluative factors to be used in

---

<sup>1095</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 34.

<sup>1096</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 34.

<sup>1097</sup> Tierney and Simon, "What's Still Wrong with Rubrics."

<sup>1098</sup> Popham, "What's Wrong--and What's Right--with Rubrics." 75.

<sup>1099</sup> Popham, Test Better, Teach Better. 99.

<sup>1100</sup> Popham, Test Better, Teach Better. 99.

<sup>1101</sup> Popham, Test Better, Teach Better. 99.

appraising a student's performance. This way, teachers and students can remain attentive to a modest number of super-important evaluative criteria, rather than being overwhelmed or sidetracked."<sup>1102</sup> Once on paper, these criteria will provide stakeholders with a clear understanding of what students must demonstrate to show that they have achieved an outcome.<sup>1103</sup> They will help "students understand what is expected of them. Listing criteria also helps the program administrators know what to teach the students, and it helps the administrators to identify whether the students have met the intended outcomes."<sup>1104</sup>

The next step of rubric creation involves articulating the number and types of levels that may be demonstrated in students' performances.<sup>1105</sup> Bresciani, Zelna and Anderson offer advice with this step:

Here, you can choose a simple rating scale or begin to describe in detail what you are looking for and what various levels look like. If you get stuck here, just think about the various levels in which you have seen students' abilities reside prior to your decision to use rubric. The descriptions may come more easily with past observations of students' learning in mind. If not, ask your colleagues and even students to assist you.<sup>1106</sup>

Moskal recommends beginning with a description of the top level of performance, then turning to the lowest level of performance.<sup>1107</sup> She notes that, "The contrast

---

<sup>1102</sup> Popham, Test Better. Teach Better. 99.

<sup>1103</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 29.

<sup>1104</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 30.

<sup>1105</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 34.

<sup>1106</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 34.

<sup>1107</sup> Moskal, "Scoring Rubrics: What, When, and How?."



between the criteria for top level performance and bottom level performance is likely to suggest appropriate criteria for the middle level of performance.”<sup>1108</sup>

Researchers suggest a number of principles to adhere to in writing levels of performance descriptions. Tierney and Simon caution rubric developers to be consistent when writing descriptions across performance levels.<sup>1109</sup> They write, “Consistency in performance criteria can basically be viewed as the reference to the same attributes in the descriptors across the levels of achievement.”<sup>1110</sup>

Farmer suggests a way of dividing levels of performance. She writes, “The difference between levels of competence often lies in degree or thoroughness, such as ‘wide variety’ versus ‘diverse’ or ‘some’ versus ‘many.’ These key words act as critical features to distinguish one level of competency from another.”<sup>1111</sup>

Tierney and Simon cite other researchers and suggest that the three measurement scales most commonly used are amount, frequency, and intensity.<sup>1112</sup> They also point out that descriptive language communicates most effectively to students and teachers.<sup>1113</sup> Other researchers emphasize the importance of using descriptive, rather than judgmental, language to describe levels of student performance.<sup>1114</sup>

---

<sup>1108</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1109</sup> Tierney and Simon, "What's Still Wrong with Rubrics."

<sup>1110</sup> Tierney and Simon, "What's Still Wrong with Rubrics."

<sup>1111</sup> Farmer, "Authentic Assessment of Information Literacy."

<sup>1112</sup> Tierney and Simon, "What's Still Wrong with Rubrics."

<sup>1113</sup> Tierney and Simon, "What's Still Wrong with Rubrics."

<sup>1114</sup> Moskal, "Scoring Rubrics: What, When, and How?."

Several researchers weigh in on the number of performance levels that should be included in a rubric. Arter and McTighe acknowledge that there are no rules governing this, but they suggest considering the nature of the performance and the overriding purpose of assessment.<sup>1115</sup> They suggest asking a number of questions, including “What is the likely range of qualitatively different degrees of understanding, proficiency, or quality in the product or performance?”<sup>1116</sup> Arter and McTighe recommend enough levels to differentiate levels of quality, but not so many that teachers and students are confused.<sup>1117</sup> They also present arguments both for and against an odd number of levels, noting that odd numbers may cause raters to tend toward the mean, but that middle numbers give raters a way to present a balance between strengths and weaknesses in student performance.<sup>1118</sup> Popham suggests that rubrics with three to five levels are best for teachers and students.<sup>1119</sup> Bresciani, Zelna and Anderson suggest two to four, and they note that it is possible that most people “can only differentiate three levels of anything—‘it is not there,’ ‘sort of there,’ and ‘all there.’”<sup>1120</sup> Stevens and Levi also recommend starting with three levels.<sup>1121</sup> Moskal reminds rubric developers that it is preferable to have “a few meaningful

---

<sup>1115</sup> Arter and McTighe, Scoring Rubrics. 29.

<sup>1116</sup> Arter and McTighe, Scoring Rubrics. 31.

<sup>1117</sup> Arter and McTighe, Scoring Rubrics. 31.

<sup>1118</sup> Arter and McTighe, Scoring Rubrics. 31.

<sup>1119</sup> Popham, "What's Wrong--and What's Right--with Rubrics." 75.

<sup>1120</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 34.

<sup>1121</sup> Stevens and Levi, Introduction to Rubrics. 8.

score categories than to have many score categories that are difficult or impossible to distinguish".<sup>1122</sup> Popham agrees, and notes that in general, short rubrics are better.<sup>1123</sup> Callison offers names for up to four levels of performance descriptions. They are "beginning," "developing," "accomplished," and "exemplary."<sup>1124</sup> Huba and Freed also offer naming schemes for performance descriptions.<sup>1125</sup>

One last area of determination for the levels of performance included in an assessment rubric is the issue of numerical weights. Moskal notes that rubrics give traditional numerical scores meaning:

The assignment of numerical weights to sub-skills within a process is another evaluation technique that may be used to determine the extent to which given criteria has been met. Numerical values, however, do not provide students with an indication as to how to improve their performance.... Scoring rubrics respond to this concern by providing descriptions at each level as to what is expected. These descriptions assist the students in understanding why they received the score that they did and what they need to do to improve their future performances.<sup>1126</sup>

Callison acknowledges that adding numerical weights to rubric performance levels can make grading easier, but cautions that adding numbers to a rubric does not replace detailed descriptions of levels of student performance.<sup>1127</sup> He states:

---

<sup>1122</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1123</sup> Popham, "What's Wrong--and What's Right--with Rubrics." 75.

<sup>1124</sup> Callison, "Rubrics." 35.

<sup>1125</sup> Mary E. Huba and Jann E. Freed, Learner-Centered Assessment on College Campuses (Boston: Allyn and Bacon, 2000). 180.

<sup>1126</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1127</sup> Callison, "Rubrics." 34.

Some rubric models allow for points at each level so the point totals can be translated into standard grades, if necessary.... Such numeric guides, however, should not take away from the purpose and value of rubrics: to write in clear statements those performances that are not acceptable and those that are acceptable as measured against the expected standard performance for the task and age group.<sup>1128</sup>

After determining outcomes, criteria, and levels of performance descriptions, the next step is to use the rubric to norm assessors.<sup>1129</sup> The purpose of this step is “to get everyone on the same page as to the meaning of the criteria and as to the identification of the varying levels that they may see students demonstrate.”<sup>1130</sup> Bresciani, Zelna and Anderson point out that when teams work together to develop a rubric, norming may begin much earlier in the development process “because there is a debate and discussion about what the criteria should be and what they look like when a student has mastered them.”<sup>1131</sup> However, when educators develop rubrics independently and then share them with others, the norming process requires special attention. First, the rubric users need to be reminded of the outcome the rubric assesses and how the evidence of learning is collected.<sup>1132</sup> Through the norming process, future users of the rubric must come to understand all of the dimensions, including criteria and levels of performance, of the rubric.<sup>1133</sup>

---

<sup>1128</sup> Callison, "Rubrics." 34.

<sup>1129</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 35.

<sup>1130</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 35.

<sup>1131</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 35.

<sup>1132</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 36.

<sup>1133</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 36.

In order to norm a group of raters to use a rubric, it is important to practice using sample evidence of student learning.<sup>1134</sup> Tierney and Simon suggest that rubrics should be accompanied by examples of student work and believe that sharing these examples with raters helps decrease variability among scores assigned by raters.<sup>1135</sup> Such examples of student work, that are pre-scored and selected to highlight the nuances of the rubric, are known as anchor papers.<sup>1136</sup> Hess states that anchor papers provide “a set of clear examples of student responses that illustrate each score point of a rubric. Having tangible examples of student work considerably enhances the reliability of subsequent scoring by classroom teachers. In addition, anchored performances can serve as models for teachers and students.”<sup>1137</sup> By providing anchor papers, “everyone knows what a ‘4’ looks like in comparison to a ‘3’ or ‘2’ or ‘1.’”<sup>1138</sup>

Two researchers report the steps of the norming process. Hess shares a four-step process for anchoring student responses for rubric assessment.<sup>1139</sup> In the first step, raters read student responses without trying to score them. Then they read the rubric. Third, in small groups, raters score a student response as a “clear example” of a level, a “possible example,” or indicate that the small group

---

<sup>1134</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 36.

<sup>1135</sup> Tierney and Simon, "What's Still Wrong with Rubrics."

<sup>1136</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1137</sup> Stephen K. Hess, "Validating and Anchoring Classroom Performance Tasks," A Handbook for Student Performance in an Era of Restructuring, eds. Robert E. Blum and Judith A. Arter (Alexandria, Virginia: Association for Supervision and Curriculum Development, 1996). V-9:2.

<sup>1138</sup> Farmer, "Authentic Assessment of Information Literacy."

<sup>1139</sup> Hess, "Validating and Anchoring Classroom Performance Tasks." V-9:3-4.

was “split without consensus.” Finally, raters focus on the clear examples of rubric levels, explaining in writing the relationship between characteristics of the student response and the corresponding rubric level. These written explanations are then shared with the large group and later prove useful to other teachers in using and interpreting the rubric in their classrooms.<sup>1140</sup> In Baker, Abedi, Linn, and Niemi’s study, raters are familiarized with the rubric and practice scoring pre-scored anchor papers over a three-hour period.<sup>1141</sup> In this study, verifying that trainers could produce reliable scores required an additional 45 minutes. Finally, raters are permitted to score student responses, but consistency is checked by inserting previously scored papers into each rater’s set “to assure that raters were maintaining appropriate score-point definitions.”<sup>1142</sup> According to these researchers, raters also scored a “randomly selected subsample of seventeen papers twice to permit analysis of intrarater reliability.”<sup>1143</sup>

Maki suggests a model six-step process to norm raters and “calibrate” them to ensure interrater reliability.<sup>1144</sup> First, she recommends asking raters to score a set of student samples independently. The set of student samples should reflect the full range of student work. Then, she suggests bringing raters together to discuss their responses and identify areas where scores converge or diverge. Next, raters should discuss and reconcile inconsistent scores. At this

---

<sup>1140</sup> Hess, “Validating and Anchoring Classroom Performance Tasks.” V-9:3-4.

<sup>1141</sup> Baker, Abedi, Linn and Niemi, “Dimensionality and Generalizability.” 200.

<sup>1142</sup> Baker, Abedi, Linn and Niemi, “Dimensionality and Generalizability.” 200.

<sup>1143</sup> Baker, Abedi, Linn and Niemi, “Dimensionality and Generalizability.” 200.

<sup>1144</sup> Bernier, “Making Yourself Indispensable.” 126-127.

point, raters should independently score a new set of student responses.

Afterwards, Maki recommends that raters repeat the process of identifying and discussing inconsistent and consistent scores across the group and then reconciling divergent scores. Maki suggest that this process be repeated until raters reach consensus in applying the rubric to student samples.<sup>1145</sup>

After norming the rater group using the rubric, the next step is to pilot the rubric on sample work.<sup>1146</sup> Bresciani, Zelna and Anderson recommend testing the rubric on a pilot sample of student work, previous student work, or on “colleagues willing to role-play as students.”<sup>1147</sup>

Once the pilot test is completed, the next and final step is to revise the rubric. When revising a rubric for assessment, the focus should be on making the rubric more meaningful to its users.<sup>1148</sup> Bresciani, Zelna and Anderson reveal that “we refine our rubrics every year because the students help us to better communicate the criteria and the levels of expected performance. There may be one day when we do not have to further refine our rubrics, but we have not yet reached that point.”<sup>1149</sup>

After all the steps of developing the rubric are complete, Bresciani, Zelna and Anderson recommend documenting findings, adding the rubric to a

---

<sup>1145</sup> Bernier, "Making Yourself Indispensable." 127.

<sup>1146</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 36.

<sup>1147</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 35.

<sup>1148</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 36.

<sup>1149</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 35.

program's overall assessment plan, and making the rubric public.<sup>1150</sup> They acknowledge the latter as, "a controversial suggestion. Some evaluators believe that the rubric should not be made public for it appears to be 'teaching to the test.' Others argue that making the rubric public means that students can better take responsibility for what is expected of them and that program administrators become more accountable for what is expected of them."<sup>1151</sup> Moskal and Leydens suggest, "Whenever possible, the scoring rubric should be shared with the students in advance to allow students the opportunity to construct the responses with the intention of providing convincing evidence that they have met the criteria."<sup>1152</sup> Bresciani, Zelna and Anderson also reason that making rubrics public does not, in fact, teach to the test. They write:

A rubric describes what the demonstrated learning and development looks like. It does not mean that the learning and development cannot be more than what is stated on the rubric. If rubrics are written well with developmental theory in mind, rubrics can be invaluable to everyone understanding the interrelatedness of learning and development. Thus, we highly recommend making rubrics public.<sup>1153</sup>

### **Documentation of Rubric Assessment Results**

Oftentimes, assessment is approached with enthusiasm and good intentions. Many times, that same energy is not applied to the documentation stage of the assessment process. Bresciani, Zelna and Anderson note, "Many

---

<sup>1150</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 36.

<sup>1151</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 36.

<sup>1152</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1153</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 36.



cocurricular professionals and faculty are still not documenting their [assessment] work.”<sup>1154</sup> They find that assessment is not always approached systematically, and so key aspects, such as documentation, can be overlooked.<sup>1155</sup> When documentation is omitted, problems with accountability ensue. “Thus, when individuals attempt to demonstrate that they have engaged in assessment by articulating outcomes, gathering data to measure those outcomes, and making decisions for continuous improvement, there is no evidence to reveal.”<sup>1156</sup> Fear may also interfere with documentation. Bresciani, Zelna and Anderson write, “It is one thing to look at data and make decisions, and quite another to commit to writing down one’s interpretation of the data.”<sup>1157</sup> They confess:

The first time we set out to write down our first assessment finding, we literally felt afraid. For some reason, we felt that writing the results down made us more vulnerable, as if someone would see our inner faults and blame us for being a bad teacher or program administrator. Now, the documenting comes with much less fear and much more liberation as we find this step helpful in simply reminding us of what we need to change the next go around and exactly why.<sup>1158</sup>

There are, however, compelling reasons to document assessment efforts, including both accountability and learning. All of higher education needs to respond to increased demands to account for the learning it produces.<sup>1159</sup>

Educators involved in assessment are responsible for demonstrating to their

---

<sup>1154</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 80.

<sup>1155</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 80.

<sup>1156</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 80.

<sup>1157</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 81.

<sup>1158</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 82.

<sup>1159</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 80.

constituents that programs are effective<sup>1160</sup> and that information gained through assessment processes is used to make improvements to further increase program effectiveness.<sup>1161</sup>

Documenting assessment also provides an opportunity for assessors to learn about both their programs and student learning.<sup>1162</sup> Bresciani, Zelna, and Anderson describe the learning they gain through documenting assessment processes. They write, “We have learned more about our programs and courses because we had to document the learning process. We believe that we need to go through the process of documenting our own assessment work for ourselves, as well as the direct benefactors of our programs and classes.”<sup>1163</sup> They also state that recording decisions made based on assessment data “serve[d] as helpful reminders to ourselves and others about why we made the decision(s) we did.”<sup>1164</sup>

A number of principles should be followed in documenting the assessment of student learning in general and rubric assessment in particular. First, the documentation of assessment results should maintain a focus on student learning outcomes. Gratch Lindauer states that results of the assessment of library instruction must be reported in terms of student outcomes, and be free

---

<sup>1160</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 80.

<sup>1161</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 82.

<sup>1162</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 82.

<sup>1163</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 80-81.

<sup>1164</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 82.

from library jargon.<sup>1165</sup> Focusing on outcomes helps assessors to see how well or poorly the method of collecting assessment data matches the outcome, to determine to what extent students were or were not learning, and to make decisions for continuous improvement.<sup>1166</sup> It also emphasizes accountability and expresses data in the terms that are most important to the stakeholder community.<sup>1167</sup>

When documenting and reporting assessment results, several strategies can be used to ensure stakeholders' understanding. Wiggins states that assessment results should be "reported and used so that all customers for the data are satisfied."<sup>1168</sup> Sometimes, this requires formatting assessment data differently for different categories of stakeholders. One stakeholder group may value different results than another. Bresciani, Zelna and Anderson suggest that assessors "identify the values of your constituents and find out how your constituents prefer to see data and reports. Oftentimes you will need to revise your presentation for each group of your constituents.... In many cases...constituents have varying values and the data you present to them must reflect their values as well as be presented in a manner that resonates with

---

<sup>1165</sup> Gratch Lindauer, "Defining and Measuring the Library's Impact."

<sup>1166</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 81-82.

<sup>1167</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 15.

<sup>1168</sup> Wiggins, "Creating Tests Worth Taking." V-6:3.

them.”<sup>1169</sup> They also note that some groups prefer empirical data or rich qualitative data.<sup>1170</sup>

When preparing documents to report assessment results to wider, non-technical constituent groups, it is important to focus on the needs of these constituent groups. MacColl and White remind assessors to “focus on the needs of the audience” by making reports “accessible and understandable to people with vested interests in the educational process.”<sup>1171</sup> They recommend focusing on accessibility and readability so that stakeholders can grasp assessment results quickly.<sup>1172</sup> MacColl and White encourage report writers to consider that theories and best practices of educational effectiveness do not appear in publications that some stakeholders are likely to have access to.<sup>1173</sup> Therefore, when applicable, the basic ideas of these theories and practices will need to be included in assessment reports. They also point out that by addressing problems of organization, terminology, and statistical representation, writers can make assessment reports much more readable. MacColl and White recommend avoiding typical scholarly writing conventions such as listing major findings only at the end of publications.<sup>1174</sup> They also suggest using plain language and

---

<sup>1169</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 15.

<sup>1170</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 27.

<sup>1171</sup> Gail S. MacColl and Kathleen D. White, "Communicating Educational Research Data to General, Nonresearcher Audiences," Practical Assessment, Research, and Evaluation 6.7 (1998).

<sup>1172</sup> MacColl and White, "Communicating Educational Research Data."

<sup>1173</sup> MacColl and White, "Communicating Educational Research Data."

<sup>1174</sup> MacColl and White, "Communicating Educational Research Data."

relying on simple tables rather than complex statistical charts to convey assessment results.<sup>1175</sup> Finally, Bresciani, Zelna and Anderson advocate releasing assessment results according to a strategic timeline that considers political climate, reporting channels, and opportunities for action.<sup>1176</sup>

### **Reporting of Rubric Assessment Results**

Reporting results from a rubric assessment is similar to reporting results from any other type of assessment; however, employing a few useful techniques can make reporting rubric assessment results easier and more accessible to constituents. Some educators shy away from using rubrics as an assessment method because they believe that the qualitative information usually associated with rubrics will not provide them with data that their stakeholders will appreciate.<sup>1177</sup> Indeed, rubrics do generate rich, detailed qualitative information that is invaluable for improving programs.<sup>1178</sup> But rubrics can also generate the holistic or quantitative information that constituents may prefer. For example, assessors can report where students fell into each section on a rubric, or summarize the categories students fell into and give examples of those performances.<sup>1179</sup> If numerical data is required, reports may include percentages of students that were scored at each level for each criteria, or each cell on a

---

<sup>1175</sup> MacColl and White, "Communicating Educational Research Data."

<sup>1176</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 15.

<sup>1177</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 37.

<sup>1178</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 37.

<sup>1179</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 37.

rubric can be given a number value and students scored can be totaled for a raw score to use in quantitative analysis<sup>1180</sup> or placed on a grading or percent scale.<sup>1181</sup> Clearly, there are a number of ways to match rubric assessment results to stakeholder needs and expectations.

### **Rubrics of Information Literacy Skills**

In the area of information literacy assessment in higher education, rubrics are far outnumbered by surveys, questionnaires, and tests. Yet, a few exist, and "it is always wise to look at assessments that others have already constructed."<sup>1182</sup> The following section describes the rubrics found in the information literacy instruction literature.

In an 2001 article, D'Angelo reports the use of a rubric employed to assess the results of a project designed to teach career research skills to students.<sup>1183</sup> In this project, students researched career options, completed library instruction sessions, wrote career reports using library resources, and took a short test.<sup>1184</sup> The career reports were assessed using a short checklist rubric (see Figure 4.1). According to D'Angelo, using a rubric yields more

---

<sup>1180</sup> Colton, Gao, Harris, Kolen, Martinovich-Barhite, Wang and Welch, Reliability Issues with Performance Assessments. 7.

<sup>1181</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 37.

<sup>1182</sup> Williams, "Creativity in Assessment." 332.

<sup>1183</sup> Barbara J. D'Angelo, "Integrating and Assessing Information Competencies in a Gateway Course," Reference Services Review 29.4 (2001).

<sup>1184</sup> D'Angelo, "Integrating and Assessing Information Competencies."

### **Adult career development, bibliographic research project, evaluation rubric**

1. The student identified an information need: (either through a thesis statement or through a clearly focused essay)  
☐ Yes ☐ No
2. The student used the correct number of sources required by the assignment:  
☐ Yes ☐ No
3. The student used a variety of sources (books, articles, Internet):  
☐ Yes ☐ No
4. The student used a variety of tools (catalogs, indexes, Internet) to locate information:  
☐ Yes ☐ No ☐ Not evident  
(yes=2 or more)
5. The student used discipline- or career-specific resources, if appropriate (trade journals, professional societies):  
☐ Yes ☐ No
6. The student used sources that are relevant to the topic:  
☐ Yes ☐ No ☐ Partially (#?)
7. The student used sources that are current, if appropriate:  
☐ Yes ☐ No ☐ Partially
8. The student integrated information from all sources into the text of the essay:  
☐ Yes ☐ No ☐ Partially
9. The student identified a continued information need and search strategy:  
☐ Yes ☐ No
10. Additional comments:

*Figure 4.1. D'Angelo's Information Literacy Rubric*

significant information about students' adoption of concepts and practices than testing.<sup>1185</sup> For example, this rubric assessment showed that students tended to use only one type of source in their research. Based on the information revealed by the rubric assessment, instructors and librarians made several changes for the following year: they improved the directions for the writing assignment and included modeling and group learning in the library instruction sessions.<sup>1186</sup> The next year, students' work improved. Their reports included a wider variety of sources, and they demonstrated better search strategy. Unfortunately, in the third year of the project the rubric was dropped from the assessment process due to time constraints and later the librarian was re-assigned to other duties.<sup>1187</sup> It should be noted that the rubric described in this project exhibits a number of weaknesses. For instance, the rubric takes the form of a checklist, and therefore lacks the detail present in a full model rubric. It also includes some vague wording, such as "clearly focused", "variety of sources", and "current", that are undefined. Still, because the rubric assessment results were used to make improvements to instruction, this is a positive example of rubric use in information literacy instruction.

In a 2002 publication, Merz and Mark supply a collection of examples of library instruction assessment.<sup>1188</sup> The collection includes a few sample rubrics (see Figures 4.2-4.5). However, no accompanying information about the use or

---

<sup>1185</sup> D'Angelo, "Integrating and Assessing Information Competencies."

<sup>1186</sup> D'Angelo, "Integrating and Assessing Information Competencies."

<sup>1187</sup> D'Angelo, "Integrating and Assessing Information Competencies."

<sup>1188</sup> Merz and Mark, Clip Note #32. 87-93.



contexts of the rubrics is provided, so it is difficult to determine the impact of these rubrics on library instruction assessment.

Perhaps the most well known rubric in information literacy assessment at the college level is Rockman's rubric of information competence, published in 2002. Rockman uses a rubric (1) to define the criteria for successful student learning of information competence principles, (2) to align content, instruction, and assessment to promote learning, (3) to facilitate the evaluation of student work using common standards, and (4) to promote collaboration between faculty and librarians.<sup>1189</sup> Rockman's rubric (see Figure 4.6) is holistic and attempts to summarize all of the Information Literacy Competency Standards for Higher Education into one rubric. As a result, the rubric is useful as a general guide for conceptualizing information literacy, but it would be difficult to use the rubric to assess student work. Because of the rubric's broad scope, some wording is quite vague. For example, some instructors may not be able to differentiate "extensive information sources" from "a sufficient number of information sources." Also, too many different performance concepts are included in each cell of the rubric and, in some cases, the performance descriptions are not consistent across criteria. For instance, the beginning level of the first criteria focuses only on a research question, while the second and third levels include details about key concepts and information sources.

---

<sup>1189</sup> Rockman, Rubrics for Assessing Information Competence.

Rubric for Assessing Research Papers				
ACRL Information Literacy Competency Standard #1				
*Variety of sources (1.2.c & d)	all from one source type	mix of 2 source types	mix of more than 2 source types	
Sources satisfy research/information need	severely unbalanced (most important literature is missing)	adequate (viewpoints limited but adequate )	balanced (good representation of viewpoints)	comprehensive
ACRL Information Literacy Competency Standard #3				
Timeliness of sources (3.2.a)	inappropriate dates for topic	mix of appropriate and inappropriate	appropriate dates for all materials	
Authority/ Reliability of sources (3.2.a)	all inappropriate authorities	mix of authoritative and non-author.	all reliable authorities	
ACRL Information Literacy Competency Standard #4				
Use of reference(s) to evaluate or illustrate specific points (4.1.c; 4.3; 3.1)	doesn't use references OR quotes or references don't seem to be serving any purpose—are just "stuck in"	quotes or references serve a purpose but are generally not well-used	uses references or quotes: -for background information -to support student's thesis -as support for a specific point	
Integrates quotes effectively (style: block quotes, signal phrases, etc.) (4.3; 3.1)	not integrated well	some quotes effectively integrated (some not)	most quotes effectively integrated	sophisticated use of quotes
Citing/ documenting materials in text (4.1.c)	many errors (did not cite accurately or neglected to cite source)	most are correct but minor errors are numerous	all entries conform to required style with few punctuation errors	
ACRL Information Literacy Competency Standard #5				
Bibliography/works cited: <i>one style</i> (5.3.a)	many errors	most entries conform to style; minor errors are numerous	all entries conform to required style with few punctuation errors	
Bibliog/works cited: <i>complete info.</i> (5.3.a)	incomplete	most are complete; some missing info.	information is complete	
Are all sources cited in text listed on Works Cited page or bibliography? (5.3.a)	source list is incomplete	source list is complete		
Are all sources listed on Works Cited page cited in paper?	none	some	most	all
Evidence of plagiarism (uncredited sources) 5.3.a	throughout paper	some	none	
Overall Rating – Student Use of Outside Sources				
Does this student know how to use outside sources, as evidenced in this paper?	poor use of sources	some ability shown	good use	sophisticated use

\*All standards & references to performance indicators (numbers in left column of rubric) are tied to ACRL's *Information Literacy Competency Standards for Higher Education* (2000).  
 Variety of sources = books, journal articles, popular magazines, websites, etc.

Developed: Beth Mark & Lawrie Merz 4/27/01

Figure 4.2. Merz and Mark's Information Literacy Rubric #1

## Assessment of First Year Seminar Paper Sample Selection, Use, and Citation of Sources – Report to Faculty, Fall 1999

SOURCE SELECTION				
Variety of source types (e.g., books, journal or magazine articles, etc.)	21=23.3% all from one type of source	41=45.5% mix of 2 types of sources	28=31% mix of more than 2 types of sources	
Timeliness of sources (for topic)	0 inappropriate dates	5=4.5% mix of appropriate and inappropriate	85=94.4% appropriate dates for all materials	
Authority/ Reliability of sources for topic	0 all inappropriate authorities	40=44.4% mix of authoritative and non-authoritative	50=55.5% all reliable authorities	
INFORMATION USE				
Sources satisfy research/ information need	10=11% severely unbalanced (most important literature is missing)	33=36.6% adequate (viewpoints limited but adequate)	37=41% balanced (good representation of viewpoints)	10=11% comprehensive
Use of reference(s) to evaluate or illustrate specific points	2=2.2% doesn't use references OR quotes or references don't seem to be serving any purpose—are just "stuck in"	18=20% quotes or references serve a purpose but are generally not well-used	36=40% (good) 34=37.7% (better) uses references or quotes: -for background information -to support student's thesis -as support for a specific point	
Integrates quotes effectively (style: block quotes, signal phrases, etc.)	6=6.6% not integrated well	23=25.5% some quotes effectively integrated (some not)	35=38.8% most quotes effectively integrated	26=28.8% sophisticated use of quotes
CITATION				
Citing/ documenting materials in text	18=20% many errors (did not cite accurately or neglected to cite source)	21=23.3% most are correct but minor errors are numerous	51=56.6% all entries conform to required style with few punctuation errors	
Bibliography/works cited: one style	14=15.5% many errors	30=33.3% most entries conform to style; minor errors are numerous	46=51% all entries conform to required style with few punctuation errors	
Bibliog/works cited: complete information	4=4.4% incomplete	40=44.4% most are complete; some missing info.	46=51% information is complete	
Are all sources cited in text listed on Works Cited page or bibliography?	15=16.6% source list is incomplete	75=83.3% source list is complete		
Are all sources in Works Cited list cited in paper?	1=1.1% none	9=10% some	19=21% most	61= 67.7% all
Evidence of plagiarism or uncredited sources	3=3.3% throughout paper	31=34.4% some	56= 62% none	
OVERALL RATING – STUDENT USE OF OUTSIDE SOURCES				
Overall rating: Does student know how to use outside sources, as evidenced in this paper?	2=2.2% poor use of sources	24=26.6% some ability shown	47=52% good use	17=18.8% sophisticated use

Developed by Beth Mark & Lawrie Merz 4/27/01

Figure 4.3. Merz and Mark's Information Literacy Rubric #2

Research Project Reference Rating (RPRR) Bibliography Assessment Rubric			
Description	Score (Circle)		
1) Student provides <b>justification for selecting sources</b> or how they will be relevant to the project (annotated bibliography or paragraph in project; 1 = yes; 0 = no)	1	0	
2) <b>MLA or APA Citation Style Format</b> (for the references)			
a) <b>Order</b> of entries (alphabetical by author or title if no author; 1 = yes; 0 = no)	1	0	
b) <b>Spacing</b> , double or single (1 = correct; 0 = incorrect)	1	0	
c) <b>Indentation</b> (1 = correct; 0 = incorrect)	1	0	
d) <b>Punctuation</b> (1 = correct; 0 = incorrect)	1	0	
e) <b>Capitalization</b> (1 = correct; 0 = incorrect)	1	0	
f) <b>Spelling, Abbreviations</b> (1 = correct; 0 = incorrect)	1	0	
g) <b>Italics/Underlining</b> (1 = correct; 0 = incorrect)	1	0	
h) <b>Names</b> (1 = correct; 0 = incorrect)	1	0	
i) <b>Dates</b> (1 = correct; 0 = incorrect)	1	0	
j) <b>Volume, Issue, Number, Pagination</b> (1 = correct; 0 = incorrect)	1	0	
k) <b>Uniform Resource Locators</b> (1 = correct; 0 = incorrect)	1	0	
l) <b>Order within entries</b> (1 = correct; 0 = incorrect)	1	0	
m) <b>Information</b> sufficient to locate source (2 = complete; 1 = some errors; 0 = inadequate)	2	1	0
3) <b>Source Balance</b>			
a) Adequate <b>number</b> to cover topic (2 = 4 or more unique; 1 = 1 to 3)	2	1	0
b) Includes <b>key primary</b> sources (2 = 2 or more; 1 = at least one; 0 = none)	2	1	0
c) Includes <b>key secondary</b> sources (2 = 2 or more; 1 = at least one; 0 = none)	2	1	0
d) <b>Variety</b> (at least one from each of the following categories; count each cite once)			
i. Background (encyclopedia, dictionary, yearbook, reference)	1	0	
ii. Book	1	0	
iii. Government document or .gov web site	1	0	
iv. Scholarly journal article	1	0	
v. Popular or magazine article	1	0	
vi. Newspaper, wire service, radio, TV, or web cast article	1	0	
vii. Web site	1	0	
viii. Interview, speech, lecture, or other primary source	1	0	
ix. Other appropriate source	1	0	
4) <b>Quality of Sources for the Selected Topic</b>			
a) <b>Currency</b> (rapidly changing fields require recent materials; historical topics require a broad spectrum; 0 = older or inappropriate sources; dates not provided; 1 = a mix of older and recent sources; 2 = all within the last 10 years or appropriate to the topic)	2	1	0
b) <b>Authority</b> of author and sponsor (produced by a government agency, higher education institution, or other reputable organization with expertise on the topic; 0 = all inappropriate authorities; 1 = a mix of authoritative and non-authoritative; 2 = all reasonable authorities)	2	1	0
c) <b>Objectivity and Accuracy</b> (0=all biased, inaccurate, or commercial sources; 1=a mix of objective and biased sources; 2=all objective sources projecting a factual, informational approach)	2	1	0
5) <b>Coverage, Scope, and Depth of Sources Cited</b>			
a) Relevant to the General <b>Field</b>	2	1	0
b) Relevant to the Specific <b>Topic</b> (the right sources to obtain information on this topic)	2	1	0
c) <b>Appropriate</b> for an academic paper (include audience; overall assessment)	2	1	0
<b>TOTAL SCORE</b>			

Figure 4.4. Merz and Mark's Information Literacy Rubric #3

**Rating Scale***SSCI 221**and Social Sciences Majors*

Student Name:

Evaluator:

Date:

---

A hypothetical research topic will be assigned to the student with instructions to devise and demonstrate a successful search strategy. Topics will be selected by the Instruction Librarian from *10,000 Ideas for Term Papers, Projects and Reports, 3rd ed.* by Kathryn Lamm [R 808.02 L188te3 1991] located in the Reference Room. The student will:

(Evaluator will circle most appropriate rating for activity)

1. Ascertain the subject field of the topic and the discipline into which the subject field fits.
  3. The subject field and the discipline are ascertained
  2. Either subject field or the discipline is ascertained
  1. Neither subject field nor the discipline are ascertained
2. Determine appropriate Subject Heading/Term(s) that best describe the concept(s) of the topic.
  3. Library of Congress Subject Heading/Term(s) descriptive of the concept(s) are used
  2. Invalid Keyword(s) -- non-Subject Headings -- somewhat descriptive of the concepts are used
  1. Invalid, nondescriptive Subject Headings or Keywords are used
3. Select appropriate/discipline-specific access tool.
  3. Selects catalog and discipline-specific electronic (or print) index
  2. Selects only a general index or only the catalog
  1. Selects neither a discipline-specific index nor catalog
4. [Obtains] one factual, objective source within the Library's collection.
  3. Obtains a factual, objective source within the Library's collection
  2. Obtains a factual source outside the Library's collection (Website outside library vendors)
  1. Obtains a nonfactual, subjective source within or outside the Library's collection
5. [Obtains] two recent analytical sources (either on-site or Web-based vendor).

**Source 1:**

  3. Obtains recent analytical source (either on-site or Web-based vendor)
  2. Obtains non-analytical source
  1. Obtains no source relevant to the topic
6. [Obtains] two recent analytical sources (either on-site or Web-based vendor).

**Source 2:**

  3. Obtains recent analytical source (either on-site or Web-based vendor)
  2. Obtains non-analytical source
  1. Obtains no source relevant to the topic

7. Explains briefly how the two analytical sources relate to the topic and to each other.
  3. Explains briefly how the two analytical sources relate to the topic and to each other
  2. Explains how two (or one) analytical source(s) relate either to the topic or to each other
  1. Cannot explain how any acquired source relates to the topic or to each other
8. Selects appropriate access tool to obtain one relevant Website.
  3. Selects appropriate access tool to obtain one relevant Website
  2. Selects appropriate access tool but fails to obtain relevant Website
  1. Selects inappropriate access tool and fails to obtain relevant Website
9. Evaluates acquired sources according to evaluation criteria of accuracy, authority, currency
  3. Successfully evaluates acquired sources according to all evaluation criteria
  2. Partially evaluates all (or only some) acquired sources
  1. Cannot evaluate acquired sources according to any criteria
10. Demonstrates knowledge of proper citation format for discipline and type of resource
  3. Successfully employs proper citation format for sources acquired
  2. Employs proper citation format for some (or all) sources with aid of style manual
  1. Cannot cite sources in proper format and does not seek assistance in style manual

---

**Numeric Summation:** \_\_\_\_\_

**26 - 30** Student demonstrates competency in the academic library research process.

**20 - 25** Student has grasp of overall concepts and is familiar with the research process, but would benefit from further practice with strategy formation and resource familiarization.

**15 - 19** Student has inadequate familiarity with the research process and library resources; cannot proceed with demonstration without prompting. Re-orientation to resources and one-on-one practice sessions are indicated.

**10 - 14** Student demonstrates no familiarity with the research process and little, if any, familiarization with the library's resources. Course should be repeated or an Individual Learning Plan should be developed in collaboration with the student's academic advisor.

**Descriptive Summation and Comments:**

Source consulted: *Classroom Assessment*. Airasian, Peter W. New York: McGraw-Hill, 1991. See Chapter 7: "Performance Assessment".

[Developer/Contact: Susan M. Foster-Harper (sfoster@ucwv.edu)]

8/30/2000

[<http://www.uchaswv.edu/library/instruct/rating.html>]

*Figure 4.5. Merz and Mark's Information Literacy Rubric #4*

<b>ACRL Standard</b>	<b>Beginning</b>	<b>Proficient</b>	<b>Advanced</b>
<b>1. Determine the Extent of the Information Needed</b>	Student is unable to effectively formulate a research question based on an information need.	Student can formulate a question that is focused and clear. Student identifies concepts related to the topic, and can find a sufficient number of information resources to meet the information need.	Question is focused, clear, and complete. Key concepts and terms are identified. Extensive information sources are identified in numerous potential formats.
<b>2. Access the Needed Information Effectively and Efficiently</b>	Student is unfocused and unclear about search strategy. Time is not used effectively and efficiently. Information gathered lacks relevance, quality, and balance.	Student executes an appropriate search strategy within a reasonable amount of time. Student can solve problems by finding a variety of relevant information resources, and can evaluate search effectiveness.	Student is aware and able to analyze search results, and evaluate the appropriateness of the variety of (or) multiple relevant sources of information that directly fulfill an information need for the particular discipline,
<b>3. Evaluate Information and its Sources Critically</b>	Student is unaware of criteria that might be used to judge information quality. Little effort is made to examine the information located	Student examines information using criteria such as authority, credibility, relevance, timeliness, and accuracy, and is able to make judgments about what to keep and what to discard.	Multiple and diverse sources and viewpoints of information are compared and evaluated according to specific criteria appropriate for the discipline. Student is able to match criteria to a specific information need, and can articulate how identified sources relate to the context of the discipline.
<b>4. Use Information Effectively to Accomplish a Specific Purpose</b>	Student is not aware of the information necessary to research a topic, and the types of data that would be	Student uses appropriate information to solve a problem, answer a question, write a paper, or other purposes	Student is aware of the breadth and depth of research on a topic, and is able to reflect on search strategy, synthesize and

	useful in formulating a convincing argument. Information is incomplete and does not support the intended purpose.		integrate information from a variety of sources, draw appropriate conclusions, and is able to clearly communicate ideas to others
<b>5. Understand the Economic, Legal, and Social Issues surrounding the Use of Information, and Access and Use Information Ethically and Legally</b>	Student is unclear regarding proper citation format, and/or copies and paraphrases the information and ideas of others without giving credit to authors. Student does not know how to distinguish between information that is objective and biased, and does not know the role that free access to information plays in a democratic society.	Student gives credit for works used by quoting and listing references. Student is an ethical consumer and producer of information, and understands how free access to information, and free expression, contribute to a democratic society.	Student understands and recognizes the concept of intellectual property, can defend him/herself if challenged, and can properly incorporate the ideas/published works of others into their own work building upon them. Student can articulate the value of information to a free and democratic society, and can use specific criteria to discern objectivity/fact from bias/propaganda.

*Figure 4.6. Rockman's Information Literacy Rubric*

Late in 2002, another information literacy rubric appeared in the library instruction literature. Emmons and Martin describe the use of a rubric to assess research papers written for an English course that were completed both before and after a new library instruction program was initiated.<sup>1190</sup> The rubric (see Figure 4.7) assess students' research paper citations based on relevance,

<sup>1190</sup> Mark Emmons and Wanda Martin, "Engaging Conversation: Evaluating the Contribution of Library Instruction to the Quality of Student Research," *College and Research Libraries* 63.6 (2002). 547.



credibility, and engagement demonstrated by the students' selection of sources.<sup>1191</sup> Unfortunately, the rubric is confusingly formatted, lacks detailed descriptions of performances, and in some criteria areas focuses more on quantity of performance than quality. Two areas that are clearly problematic are the performance descriptions for "authority" and "evidence" which are confusing and somewhat inconsistent. Still, by comparing the research papers written before and after the instruction sessions, librarians found that there was no difference in the total number of citations students used, the sources were no more recent, and the citations were still frequently inaccurate. They also found that students cited more scholarly journal articles but the same number of books, and students were more able to understand, interpret, and discuss information than evaluate, analyze, or synthesize it.<sup>1192</sup> Although the library instruction sessions did not make the impact the librarians hoped, the collaboration between faculty and librarians improved the way faculty viewed librarians and paved the way for future instructional partnerships.<sup>1193</sup>

In 2003, Choinski, Mark, and Murphey report the use of a rubric to score reflection papers written in a for-credit information literacy course.<sup>1194</sup> The rubric includes each outcome for the assignment and a five-point scale.<sup>1195</sup> Using this

---

<sup>1191</sup> Emmons and Martin, "Engaging Conversation." 550.

<sup>1192</sup> Emmons and Martin, "Engaging Conversation." 554-558.

<sup>1193</sup> Emmons and Martin, "Engaging Conversation." 560.

<sup>1194</sup> Choinski, Mark and Murphey, "Assessment with Rubrics." 563.

<sup>1195</sup> Choinski, Mark and Murphey, "Assessment with Rubrics." 563.

**FIGURE 2**  
**The Rubric**

**English 102 Portfolios—Research Paper Rubric**  
Fall Semester, 2001

**READER** \_\_\_\_\_

Portfolio Number: \_\_\_\_\_  
English 102 section: \_\_\_\_\_  
Fall Spring Summer  
Course Grade \_\_\_\_\_

**From Library:**

Did the class come to the library?

★ YES ★ NO

Did the class take the second optional  
lab session? ★ YES ★ NO

**DESCRIPTION**

**Format**

Total	Textbook or Reading	Reference	Book	Scholarly Journal	Popular Magazine	Web	Other

*List total number of citations followed by total number of each format.*

**Journals and Magazines**

EBSCO index	EBSCO full text	Total journals/magazines

*List number indexed in EBSCO and number available in full text. List total journals and magazines cited.*

**Time Frame**

3	2	1
Within five years	Five to fifteen years old	More than fifteen years old

*List number published in each time frame.*

**Accuracy**

3	2	1
All references in standard MLA style	Includes identifying information, with errors in format	Insufficient or incorrect information, frequent errors

*See following pages for relevance, credibility, and engagement scores.*

	Content	Currency	Level	Total
<b>Relevance</b>				
	Authority	Summary	Purpose	Total
<b>Credibility</b>				
	Evidence	Challenge	Meaning	Total
<b>Engagement</b>				

**FIGURE 2**  
**The Rubric (continued)**

**RELEVANCE/SIGNIFICANCE**

Student selects information that addresses the thesis or question and that helps analyze a problem or propose a solution

	<b>4/Excellent</b>	<b>3/Adequate</b>	<b>2/Limited</b>	<b>1/Poor</b>
<b>Content</b> Source relevant to topic	All sources clearly related to topic.	Most sources clearly related to topic.	Some sources clearly related to topic.	Sources unrelated to topic or relevance unclear.
<b>Currency</b> Time frame appropriate to topic	All sources published in appropriate time frame	Most sources published in appropriate time frame	Some sources published in appropriate time frame	No sources published in appropriate time frame; No attention at all to time
<b>Level</b> Level and variety appropriate to purpose	All sources written at level appropriate to purpose	Most sources written at level appropriate to purpose	Some sources written at level appropriate to purpose	No sources written at level appropriate to purpose

**CREDIBILITY**

Student cites authority of author, summarizes main ideas, and recognizes bias.

	<b>4/Excellent</b>	<b>3/Adequate</b>	<b>2/Limited</b>	<b>1/Poor</b>
<b>Authority</b> Recognizes author's authority	Identifies each author's credentials – and they are relevant	Identifies author's credentials	Provides limited information about authors	Provides no information about authors
<b>Summary</b> Summarizes main idea	Summarizes main idea of each source and makes explicit connection with the argument	Summarizes main idea of most sources	Summarizes main idea of some sources	Does not summarize main idea of sources
<b>Purpose</b> Acknowledges author's bias and purpose	Explicitly acknowledges the purpose or bias of each source	Implicitly acknowledges the purpose or bias of each source	Acknowledges the purpose or bias of some sources	Does not acknowledge the purpose or bias of sources

<p style="text-align: center;"><b>FIGURE 2</b> <b>The Rubric (continued)</b></p>				
<p><b>ENGAGEMENT</b> Student constructs knowledge and makes meaning of information by entering into a conversation and arguing with sources.</p>				
	<b>4/Excellent</b>	<b>3/Adequate</b>	<b>2/Limited</b>	<b>1/Poor</b>
<b>Evidence</b> Supports arguments with evidence	Elaborates and extends by supporting arguments with evidence throughout	Does not question the validity of sources.	Quotes sources without comment and without critical understanding or evaluation.	Makes unsupported assertions throughout with little or no effort to cite sources
<b>Challenge</b> Challenges ideas and recognizes more than one side to an issue	Paper as a whole acknowledges more than one perspective; criticizes own perspective	Acknowledges more than one perspective.	Acknowledges only own perspective and denies validity of other perspectives.	Does not acknowledge more than one perspective
<b>Meaning</b> Constructs meaning or presents original idea	Analyzes and synthesizes; provides new interpretations of old ideas; draws connections	Interprets and evaluates; compares and contrasts ideas; assesses value of ideas	Understands, interprets, and discusses ideas	Presents no original ideas; shows poor understanding

*Figure 4.7. Emmonds and Martin's Information Literacy Rubric*

rubric revealed that students had difficulty evaluating websites, and only two-thirds included criteria of websites suitable for academic use.<sup>1196</sup> The authors explain:

Most of the unsuccessful answers provided incomplete responses, not fully answering the question. For example, a student would say it is important to find the author, but would not mention the reason for doing so, such as establishing the institutional affiliation, credibility, or educational status of the site author. Another says to look at the "source" of the material, but didn't specify URL, author, organization, or works cited.<sup>1197</sup>

<sup>1196</sup> Choinski, Mark and Murphey, "Assessment with Rubrics." 566.

<sup>1197</sup> Choinski, Mark and Murphey, "Assessment with Rubrics." 556, 571.

Rubric assessment also revealed that students scored poorly in areas that required higher-order reasoning.<sup>1198</sup> Despite their disappointment with student performance, Choinski, Mark, and Murphey state that using rubrics for assessment “provided objective results directly related to [our outcomes].”<sup>1199</sup> They continue, “Each rubric item was evaluated across the responses giving us a clear picture of where students as a whole have the most problems. This allowed us to determine what areas should be improved the next time the class is taught.”<sup>1200</sup> The authors believe that refining their rubrics will result in greater efficiency and proclaim rubrics to be “usable and useful tool[s] to add to the assessment arsenal.”<sup>1201</sup>

Also in 2003, a collection of case studies of outcomes-based assessment used to assess library instruction was published. Although the majority of the assessments in the collection used outcomes-based tests to assess information literacy skills, a number of rubrics were included. For example, Buchanan reports the rubric assessment of a website evaluation assignment intended to measure students’ application of authority, accuracy, coverage, currency, and objectivity as criteria for web evaluation.<sup>1202</sup> Although Buchanan’s rubric is very focused on the number of times a student exhibits a behavior and the

---

<sup>1198</sup> Choinski, Mark and Murphey, "Assessment with Rubrics." 571-572.

<sup>1199</sup> Choinski, Mark and Murphey, "Assessment with Rubrics." 572.

<sup>1200</sup> Choinski, Mark and Murphey, "Assessment with Rubrics." 572.

<sup>1201</sup> Choinski, Mark and Murphey, "Assessment with Rubrics." 573.

<sup>1202</sup> Lori E. Buchanan, "Assessing Liberal Arts Classes," Assessing Student Learning Outcomes for Information Literacy Instruction in Academic Institutions, ed. Elizabeth Fuseler Avery (Chicago: Association of College and Research Libraries, 2003). 70.

performance descriptions for the “content of essay” criteria are vague (see Figure 4.8), it still yields interesting results. Buchanan learned that less than half of students could apply all five evaluation criteria.<sup>1203</sup> She also notes that students were most successful in applying coverage and currency to evaluate websites,

These are the **grading** criteria:

<b>Score Levels</b>	<b>Number of Criteria Covered</b>	<b>Information Verification</b>	<b>Content of Essay</b>
A	Applied all 5 evaluative criteria (accuracy, authority, coverage, currency, and objectivity)	Stated 2-3 possible steps to take in verifying information on Web site	Detailed, in-depth, clear descriptions of appropriately applied criteria
B	Applied 4 of 5 evaluative criteria listed above	Stated 1-2 possible steps to take in verifying information on Web site	Understandable and clear, but insufficient detail and depth
C	Applied 3 of 5 evaluative criteria listed above	Stated 1 possible step to take in verifying information on Web site	Understandable, but abbreviated and not entirely clear or simply copying some class materials
F	Applied less than 3 of the 5 evaluative criteria listed above	Stated no verification steps	Vague or nonsensical sentences, or copies class materials with no original work

*Figure 4.8. Buchanan’s Information Literacy Rubric*

and less successful using authority, accuracy, and objectivity.<sup>1204</sup>

In the same collection, Franks describes a collaborative approach to assessment of assignments given in an introductory education course. Franks recounts the work of a librarian and a course instructor who joined forces to

<sup>1203</sup> Buchanan, "Assessing Liberal Arts Classes." 72.

<sup>1204</sup> Buchanan, "Assessing Liberal Arts Classes." 72.

develop a scoring rubric for each major assignment in the course.<sup>1205</sup> These rubrics are focused on the specific tasks required to complete each assignment (see Figure 4.9) that they may not be useful in assessing other tasks. They also include vague wording. However, Franks provides some interesting reflection on the process of using rubrics for assessment. For instance, she admits that it is difficult to conceptualize criteria to include in a rubric without having examples of student work in hand.<sup>1206</sup> She reports that she “experimented with various numbers of columns (up to five) and various terms to describe the levels of competence but found that [she] was splitting hairs when [she] had more than three.”<sup>1207</sup> She also notes that the rubric saved grading time and allowed her to simplify the process of reporting her data, which she grouped by the number of students, such as “majority”, “half”, or “no”, who demonstrated a particular level of competency.<sup>1208</sup> Franks closes with a number of changes she plans to make for the next round of assessment. They included revising and improving the rubrics, giving students rubrics when the assignment is initially made, and adding requirements that students assess themselves using the rubric and review their self-assessment with a librarian.<sup>1209</sup>

---

<sup>1205</sup> Dana Franks, "Using Rubrics to Assess Information Literacy Attainment in a Community College Education Class," Assessing Student Learning Outcomes for Information Literacy Instruction in Academic Institutions, ed. Elizabeth Fuseler Avery (Chicago: Association of College and Research Libraries, 2003). 133.

<sup>1206</sup> Franks, "Using Rubrics to Assess Information Literacy." 136-137.

<sup>1207</sup> Franks, "Using Rubrics to Assess Information Literacy." 135.

<sup>1208</sup> Franks, "Using Rubrics to Assess Information Literacy." 135-136.

<sup>1209</sup> Franks, "Using Rubrics to Assess Information Literacy." 137.

A third rubric assessment in this collection is described by Gauss and Kinkema. In this study, students created webliographies, or bibliographies of websites and articles, on health topics.<sup>1210</sup> To score the webliographies, a task-focused rubric was used (see Figure 4.10).<sup>1211</sup> After using this rubric, Gauss and Kinkema plan to revise their rubric, add more “stages”,<sup>1212</sup> and require students to achieve the basic level of each phase of the rubric before allowing them to move on to the subsequent stage of the assignment.<sup>1213</sup> Gauss and Kinkema’s rubric could also be improved by making it less task specific, removing the focus on quantity, and maintaining the same number of performance descriptions across criteria.

In the same collection of assessment projects, Hutchins describes a study involving a librarian and two political science professors who collaborated to develop an annotated bibliography assignment and accompanying rubric.<sup>1214</sup> Hutchins reports that the librarian and a professor used the rubrics independently

---

<sup>1210</sup> Nancy Gauss and Kathleen Kinkema, "Webliography Assignment for a Lifetime Wellness Class," Assessing Student Learning Outcomes for Information Literacy Instruction in Academic Institutions, ed. Elizabeth Fuseler Avery (Chicago: Association of College and Research Libraries, 2003). 163.

<sup>1211</sup> Gauss and Kinkema, "Webliography Assignment." 169-171.

<sup>1212</sup> Stevens and Levi, Introduction to Rubrics. 102.

<sup>1213</sup> Gauss and Kinkema, "Webliography Assignment." 166.

<sup>1214</sup> Elizabeth O. Hutchins, "Assessing Student Learning Outcomes in Political Science Classes," Assessing Student Learning Outcomes for Information Literacy Instruction in Academic Institutions, ed. Elizabeth Fuseler Avery (Chicago: Association of College and Research Libraries, 2003). 174-175.



<b>Rubric for Scoring: <i>Great Finds Assignment</i></b> <b>Education 110</b>			
Criteria for Evaluation	3 – Exemplary	2 – Competent	1-Emergent
Students will identify at least one source that substantially increases their vocabulary surrounding their topics.	The selected vocabulary source and its annotation clearly demonstrate a substantial increase in vocabulary useful for further research on the topic.	The selected vocabulary source provides some terminology pertinent to the topic, but the annotation does not demonstrate that it substantially increases vocabulary useful for further research on the topic.	The annotation does not indicate how the source provided relevant vocabulary useful for subsequent searches.
Students will select one each: a background source, a vocabulary source, a statistical source, and a current events source related to their topic.	All four types of sources are listed and they fall under the four assigned categories.	Two or three sources are listed, and they fall under the assigned categories, or four sources are listed and they don't all fall under the assigned categories.	One or fewer sources are listed, or more are listed but they don't fall under the assigned categories.
Students will identify at least four library resources, in at least two formats, that, taken together, provide substantial treatment of their topic.	Resources are in at least two formats and, taken together, they provide substantial treatment of the topic and concepts articulated on the <i>Preliminary Preparation</i> sheet, number 5.	Resources are all in one format, or they provide essentially the same information as one another though they do treat the main topic.	There are fewer than four resources, or they are only marginally relevant to the topic.
Students will be able to describe how each of their four sources provided the information they needed.	Annotations clearly state how the sources provided the required information.	Annotations describe the sources but don't sufficiently relate them to the assigned categories.	Assignment is not annotated or annotations are limited to short, vague phrases.
Students will be able to describe the process they used to evaluate the reliability, validity, accuracy, authority, timeliness, and point of view or bias for each of these sources.	Students clearly describe the evaluative process they used in judging the reliability, validity, accuracy, authority, timeliness, and point of view or bias for each of their four sources.	Students give opinions about the reliability, validity, accuracy, authority, timeliness, and point of view or bias for each of their four sources but don't describe the evaluative process they used.	Students do not evaluate the sources.
Students will cite four sources using APA citation style.	Sources are cited correctly using APA citation style.	For each citation, three or more of the following— author, title, journal title, volume, date, pages or URL —are provided, but they are not correctly formatted.	Fewer than three elements of any citation are provided and they are not correctly formatted.

Figure 4.9. Franks' Information Literacy Rubric

Webliography Assignment Grading Rubric			
<b>Introduction</b>	<b>Basic*</b>	<b>Proficient/Advanced</b>	
Description of the topic to be addressed by the information sources	Instructor approves topic. The topic is similar to one identified in class, or of personal interest. Topic is described for uninformed reader.	Instructor approves topic. The topic is similar to one identified in class, or of personal interest. Topic is described for uninformed reader. Description includes justifications and relevant citations for why topic is timely.	
Description of the intended audience	Audience is broadly described. Statement on why audience would be interested in this webliography is lacking.	Audience is described in specific detail with identifying characteristics relevant to the topic. Includes a statement on why audience would be interested in this webliography.	
Statement on how the articles and web sites are arranged.	Arrangement of sources is described.	Arrangement of sources is described and tied to needs of the audience.	
<b>Criteria</b>	<b>Basic*</b>	<b>Proficient</b>	<b>Advanced</b>
Criteria for selecting the articles and web sites	Criteria are general and not tied to the specific audience (e.g., currency, authority of author, generally relevant to topic)	Appropriate general criteria are described, as well as 1 specific criterion relevant to the characteristics of the audience and topic.	Appropriate general criteria are described, as well as at least 2 criteria relevant to the characteristics of the audience and topic.
<b>Databases and Search Engines</b>	<b>Basic*</b>	<b>Proficient</b>	<b>Advanced</b>
List of databases and search engines used to find articles & web sites	List includes at least one periodical database and at least one search engine.	Description includes at least 3 databases and search engines selected, with at least one of each.	Description includes at least 3 databases and search engines selected, with at least one of each, and justifications for why these databases and search engines were selected for this topic
<b>Keywords, Phrases, and Search Statements</b>	<b>Basic*</b>	<b>Proficient/Advanced</b>	
Lists keywords and phrases that were successful in finding sources. Databases and search engines are listed along with the specific search statements used in searching each one.	Keywords and phrases used in searching are listed, but relevant synonyms are lacking. Databases and search engines are listed along with the specific search statements used in searching each one.	Keywords, relevant synonyms, and related terms are listed. Databases and search engines are listed along with the specific search statements used in searching each one. Appropriate search commands specific to each database or search engine are indicated.	
<b>Annotated Resources</b>	<b>Basic*</b>	<b>Proficient</b>	<b>Advanced</b>
5 articles and 5 web sites are selected based on the established criteria	50% of the sources selected meet the established criteria. Some sources repeat the same information.	Over 50% of the sources selected meet the established criteria. Each source adds some new information on the topic.	All of the sources selected meet the established criteria. Each source adds some new information on the topic.
Sources are cited in APA format.	All citations include information sufficient to identify and locate the sources. APA format is followed but 50% or more contain some errors.	All citations include information sufficient to identify and locate the sources. APA format is followed but less than 50% contain some errors.	All citations include information sufficient to identify and locate the sources. APA format is followed correctly with no error.
Abstracts or summaries	50% of the abstracts include: a brief summary of the article or site, an evaluative statement about the credibility of the author or the author's organization, and a statement about the potential value to the audience.	Over 50% of the abstracts include: a brief summary of the article or site, an evaluative statement about the credibility of the author or the author's organization, and a statement about the potential value to the audience.	All of the abstract includes: a brief summary of the article or site, an evaluative statement about the credibility of the author or the author's organization, and a statement about the potential value to the audience
<b>Writing Proficiency</b>	Over 50% of the abstracts include errors in grammar, mechanics or spelling. Sentences are somewhat mechanical or awkward. Vocabulary could be confusing or inappropriate for the audience.	50% or less of the abstracts include errors in grammar, mechanics or spelling. Sentences are constructed properly. Vocabulary is precise, interesting, and natural.	Abstracts include no errors in grammar, mechanics or spelling. Sentences are constructed properly. Writing style and vocabulary, while precise, interesting, and natural, is appropriate for the audience.
* Anything less than "Basic" will be considered "Inadequate." Absence of any of the portions will be assigned a "0."			

Figure 4.10. Gauss and Kinkema's Information Literacy Rubric

to see if interrater reliability could be established, but does not supply the results of this analysis.<sup>1215</sup> She also states that the librarian attempted to correlate rubric scores and course grades, but does not report the results of this correlation.<sup>1216</sup> However, Hutchins reflects on other impacts of rubric usage. She states that the rubric “served as a catalyst for extraordinarily fruitful conversations among library and political science faculty,” and that other political science faculty asked for permission to use the rubric in their classes.<sup>1217</sup> Students were asked to respond to the use of the rubric as well. Students reported that the specificity of the rubric was helpful, and they appreciated knowing exactly what was expected of them. On the other hand, they also noted problems with the time the rubric was distributed, and some reported the need to repeat research steps to comply with the rubric, calling it “busy work”.<sup>1218</sup> Hutchins also notes that some students, for whom English was not a first language, found the rubric difficult to understand.<sup>1219</sup> Rubric assessment resulted in a few challenges for the librarian and instructors. First, Hutchins states that it was difficult to find time to meet and discuss this collaborative approach.<sup>1220</sup> It was also time-consuming to combine political science and information literacy standards and compile rubric criteria.<sup>1221</sup>

---

<sup>1215</sup> Hutchins, "Assessing Student Learning Outcomes." 175.

<sup>1216</sup> Hutchins, "Assessing Student Learning Outcomes." 175.

<sup>1217</sup> Hutchins, "Assessing Student Learning Outcomes." 176.

<sup>1218</sup> Hutchins, "Assessing Student Learning Outcomes." 176.

<sup>1219</sup> Hutchins, "Assessing Student Learning Outcomes." 177.

<sup>1220</sup> Hutchins, "Assessing Student Learning Outcomes." 177.

<sup>1221</sup> Hutchins, "Assessing Student Learning Outcomes." 177.

Finally, the grading of student work was time-consuming, but Hutchins notes that using a scoring rubric resulted in a more consistent approach.<sup>1222</sup>

In the same collection, Kivel recounts the development of an assessment instrument used to determine whether students have reached a level of information literacy skills appropriate for a two-year college.<sup>1223</sup> While the major part of the assessment project focused on a selected-response test, there is also a performance assessment included.<sup>1224</sup> Unfortunately, the rubric used to assess the performance assessment (see Figure 4.11) “was developed only to determine whether an exam response meets the criteria of information literacy competence and it doesn’t provide a system for determine what degree of satisfactory answer has been provided.”<sup>1225</sup> In other words, the rubric focuses more on whether or not students followed directions and completed tasks rather than how well they performed. Indeed, Kivel claims creating a scoring rubric was the biggest challenge of the assessment project. He writes, “Deconstructing each activity into discreet and describable criteria and then determining the level of correctness needed to reach the bar of information competence was particularly difficult.”<sup>1226</sup> Kivel reports that future plans include rubric revision and efforts to establish the interrater reliability of the rubric.<sup>1227</sup>

---

<sup>1222</sup> Hutchins, "Assessing Student Learning Outcomes." 177.

<sup>1223</sup> Kivel, "SF Bay Area Community College." 193.

<sup>1224</sup> Kivel, "SF Bay Area Community College." 194.

<sup>1225</sup> Kivel, "SF Bay Area Community College." 194.

<sup>1226</sup> Kivel, "SF Bay Area Community College." 195

<sup>1227</sup> Kivel, "SF Bay Area Community College." 196.

**Information Competency Proficiency Exam – Performance Component**  
**SCORING RUBRIC WITH MAPPING OF LEARNER OUTCOMES**  
Version 2.3—Page 1

Outcome	Test item	Criteria to apply	What constitutes competent
1.1.4 Modifies the information need or research question to achieve a manageable focus.	B.1.a Your instructor has given you the broad subject <i>civil rights in America</i> and an assignment to write a 3- to 5-page research paper on some aspect of this topic. Narrow this subject to a manageable topic for the assignment.	B.1.a 1. Topic is narrowed by specifying time frame, or persons, or organization or group, or location, or event or incident, or some combination of these, or other similar, appropriate limiter(s) is (are) applied. AND 2. Narrowed topic is within subject assigned. AND 3. Narrowed topic is appropriate to a 3-5 page research paper.	B.1.a At least one of the narrowing techniques has been applied to the subject. AND The other two criteria are met.
1.1.3 Effectively uses background information sources to increase familiarity with the topic.	B.1.b In two or three sentences, describe the steps you took to develop your narrowed topic.	B.1.b The student describes the process used to 1. Consult additional resources, such as back-ground information sources, to develop a narrowed topic. AND/OR 2. Implement his/her own prior knowledge to develop a narrowed topic.	B.1.b What the student writes indicates that the process used meets Criterion 1 AND/OR Criterion 2
1.1.2 Formulates appropriate question(s) based on information need or research topic.	B.2 Next, take your narrowed topic and pose it as a research question you might address in this 3- to 5-page writing assignment.	B.2 1. The student composes a research question that contains a subject. AND 2. The question is a recognizable <i>type</i> of question (such as a question of fact; a posited hypothesis; some topic for comparing/contrasting; or it is a probing or investigative question; etc.). AND 3. The research question is within the narrowed topic. AND 4. The research question is appropriate to the assignment.	B.2 The research question meets all four of the criteria.

*Figure 4.11. Kivel's Information Literacy Rubric*

In a separate project included in the same 2003 collection, Knight reports the assessment of literature reviews and bibliographies of student capstone

research papers in an international studies program.<sup>1228</sup> In this study, international studies learning outcomes were mapped to the Information Literacy Competency Standards for Higher Education.<sup>1229</sup> Then, Knight used the Information Literacy Competency Standards for Higher Education to develop rubrics with three levels of competency (see Figure 4.12).<sup>1230</sup> From an outside perspective, it is difficult to understand the connections between the outcomes and performance descriptions at each level of competency in this rubric. In addition, the rubric seems to focus more on quantity than quality under the first and second standards, and the performance descriptions under the third standard are inconsistent. However, Knight does not mention any difficulties using the rubric. Rather, Knight lists deciphering incorrect citations and replicating students' research path without assigning extra work in the form of research journals as the main challenges to her assessment.<sup>1231</sup>

The 2003 collection of outcomes-based information literacy assessments includes two final descriptions of rubric approaches. First, Kobritz reports her efforts to use a rubric to assess student work in a communications capstone

---

<sup>1228</sup> Lorrie A. Knight, "Assessing Student Learning through the Analysis of Research Papers," Assessing Student Learning Outcomes for Information Literacy Instruction in Academic Institutions, ed. Elizabeth Fuseler Avery (Chicago: Association of College and Research Libraries, 2003). 202.

<sup>1229</sup> Knight, "Assessing Student Learning." 202.

<sup>1230</sup> Knight, "Assessing Student Learning." 202-203, 205-206.

<sup>1231</sup> Knight, "Assessing Student Learning." 203-204.

<p style="text-align: center;"><b>Assessment Rubric</b></p> <p>The standards, performance indicators, and outcomes are based on <i>Information Literacy Competency Standards for Higher Education</i>, Association of College and Research Libraries, January 2000.</p> <p><b>Categories:</b>      Beginner = 1      Intermediate = 2      Experienced = 3</p> <p><b>Standard 1:</b> The information-literate student determines the nature and extent of the information needed.</p> <p>    <b>Performance Indicator:</b> The information-literate student identifies a variety of types and formats of potential information sources.</p> <p>        • <b>Outcomes:</b> Identifies the value of resources in a variety of formats (books, journals, Web sites) and for different purposes (scholarly, popular)</p> <p>            o <b>Experienced:</b> At least one half of the sources will be from books published by academic presses or from scholarly journals.</p> <p>            o <b>Intermediate:</b> At least one third of the sources will be from books published by academic presses or from scholarly journals.</p> <p>            o <b>Beginner:</b> Fewer than three of the sources will be from books published by academic presses or from scholarly journals.</p> <p><b>Standard 2:</b> The information-literate student accesses information effectively and efficiently.</p> <p>    <b>Performance Indicator:</b> The information-literate student retrieves information online or in person using a variety of methods and understands that the universe of information is larger than that of the institution's library.</p> <p>        • <b>Outcomes:</b> Uses specialized online or in-person services available to retrieve information, such as ILL, full-text databases, and document delivery.</p> <p>            o <b>Experienced:</b> At least 20% of the sources were retrieved from sources other than the host library resources.</p> <p>            o <b>Intermediate:</b> At least one source was retrieved from a source other than the host library resources.</p> <p>            o <b>Beginner:</b> All sources were retrieved from the host library</p> <p><b>Standard 3:</b> The information-literate student critically evaluates information and its sources and incorporates selected information in his or her knowledge base.</p> <p>    <b>Performance Indicator:</b> The information-literate student summarizes the main ideas from the information gathered and compares new knowledge with prior knowledge to determine value added or contradictions.</p> <p>        • <b>Outcomes:</b> Identifies verbatim material that can be quoted and determines if the information satisfies the information need.</p> <p>            o <b>Experienced:</b> In the literature review, the student discusses the major authors/theories of the field of study.</p> <p>            o <b>Intermediate:</b> In the literature review, the student explains why sources were used.</p> <p>            o <b>Beginner:</b> In the literature review, the student mentions sources without evaluation of why they were selected.</p> <p><b>Standard 4:</b> The information-literate student accesses and uses information ethically and legally.</p> <p>    <b>Performance Indicator:</b> The information-literate student acknowledges the use of information sources in communicating the product or performance.</p> <p>        • <b>Outcomes:</b> Selects an appropriate documentation style and uses it consistently to cite sources.</p> <p>            o <b>Experienced:</b> The student distinguishes between online and print sources in accordance with current style sheets.</p> <p>            o <b>Intermediate:</b> The student cites sources in a consistent manner.</p> <p>            o <b>Beginner:</b> The student fails to cite sources or does so in a haphazard manner.</p> <p style="text-align: center;"><b>Example of Scoring Instrument</b></p> <table> <tr> <th>Student #</th> <th>Std. 1 Score</th> <th>Std. 2 Score</th> <th>Std. 3 Score</th> <th>Std. 4 Score</th> <th>Total Score</th> </tr> <tr> <td> </td> <td> </td> <td> </td> <td> </td> <td> </td> <td> </td> </tr> </table>						Student #	Std. 1 Score	Std. 2 Score	Std. 3 Score	Std. 4 Score	Total Score						
Student #	Std. 1 Score	Std. 2 Score	Std. 3 Score	Std. 4 Score	Total Score												

Figure 4.12. Knight's Information Literacy Rubric

course.<sup>1232</sup> Kobritz states that she chose a rubric approach to assessment in order to evaluate the higher-order thinking skills that students may display as they select resources for an assignment and explain why they felt the resources were reliable.<sup>1233</sup> Kobritz's rubric (see Figure 4.13) uses four levels of performance competency and may be difficult to use because it includes several "and/or" statements at each level of competency. The rubric could be improved by further dividing each criteria included in the rubric to eliminate the and/or statements. Still, Kobritz used the rubric to learn that students are more proficient at mechanical skills than they are at higher-order skills.<sup>1234</sup> She plans to make several changes in the future, including distributing rubrics to students earlier, providing students with a model assignment, and increasing library instruction of higher-order skills like synthesis.<sup>1235</sup>

Finally, Warmkessel describes how she collaborated with a first-year writing instructor to assess student writing using a rubric focused on students' abilities to find information and evaluate the sources they find.<sup>1236</sup> Warmkessel

---

<sup>1232</sup> Barbara Kobritz, "Information Literacy in Community College Communications Courses," Assessing Student Learning Outcomes for Information Literacy Instruction in Academic Institutions, ed. Elizabeth Fuseler Avery (Chicago: Association of College and Research, 2003). 208.

<sup>1233</sup> Kobritz, "Information Literacy in Community College." 209.

<sup>1234</sup> Kobritz, "Information Literacy in Community College." 210.

<sup>1235</sup> Kobritz, "Information Literacy in Community College." 211.

<sup>1236</sup> Marjorie M. Warmkessel, "Assessing Abilities of Freshmen to Reconcile New Knowledge with Prior Knowledge," Assessing Student Learning Outcomes for Information Literacy Instruction in Academic Institutions, ed. Elizabeth Fuseler Avery (Chicago: Association of College and Research Libraries, 2003). 250.



### Evaluation of Student Research Process

OBJECTIVE	NEEDS FURTHER INSTRUCTION	MINIMALLY ACCEPTABLE	GOOD PROGRESS	MASTERY
<p><b>STRATEGY</b></p> <p>I.4.1 Midway through the research project(C) students (A) will submit a written description of information gathered and information still needed. (B)</p> <p>III.6.3 As part of the research project (C ) students (A) will confer with experts (B).</p>	<p>Research summary lacks cohesion; sources do not address the same aspect of the topic; <i>and/or</i></p> <p>Research shows a lack of persistence in trying varying search strategies; <i>and/or</i></p> <p>Research summary does not include at least one human expert; <i>and/or</i></p> <p>Research summary indicates a lack of understanding of knowledge gaps.</p>	<p>Research summary lists a variety of types of sources, <i>and</i></p> <p>Sources selected are mostly or completely focused on the same aspect of the topic; <i>and</i></p> <p>Research summary includes at least one human expert; <i>and</i></p> <p>Research summary proposes at least one strategy for continuing the search.</p>	<p>Research summary includes a variety of opinions focused on the same aspect of the issue; <i>and</i></p> <p>Research summary demonstrates persistence in following clues from one source to another; <i>and</i></p> <p>Research summary includes at least one human expert; <i>and</i></p> <p>Research summary lists some further strategies for continuing the search.</p>	<p>Research summary demonstrates substantial effort put forth to find the seminal sources and experts on the topic; <i>and</i></p> <p>Research summary demonstrates that the student has a grasp of the contours of the issue and the major organizations and experts involved; <i>and</i></p> <p>Research summary demonstrates that the student understands remaining gaps in his or her knowledge and has a plan for closing those gaps.</p>
<p><b>EVALUATION</b></p> <p>I.2.5 As part of the research project (C) students (A) will describe each source used as primary research, secondary research or some other type of information which they can describe (B).</p> <p>I.4.2 As part of the research project (C) students (A) will submit a source sheet for each resource used with a written description of how sources were chosen and strengths and weaknesses of each. (B)</p> <p>III.2.1 III.5.1 As part of the research project (C) students (A) will submit a well-rounded selection of resources that reflect concern for accuracy, objectivity and timeliness. (B)</p>	<p>Source evaluations indicate substantial confusion as to the description of particular sources, <i>and/or</i></p> <p>Source evaluations indicate substantial confusion about authors' theses and evidence, <i>and/or</i></p> <p>Source evaluations indicate an inability to verify authors' credentials.</p>	<p>Source evaluations are inconsistent in describing types of material; <i>and/or</i></p> <p>Source evaluations are inconsistent in explaining authors' arguments and supporting evidence; <i>and/or</i></p> <p>Source evaluations ignore the author's credentials.</p>	<p>Any errors in describing sources are minor; <i>and</i></p> <p>Student clearly grasps the concept of analyzing the argument and evidence; <i>and/or</i></p> <p>Student uses any obvious external clues to reliability, such as author identification, included in the source</p>	<p>Source evaluations accurately describe each source; <i>and</i></p> <p>Source evaluations display concern for the internal logic of each piece; <i>and</i></p> <p>Source evaluations clearly demonstrate the usefulness of each source in the present context; <i>and</i></p> <p>Student uses any obvious external clues to reliability, such as author identification, included in the source.</p>

<b>SYNTHESIS</b> <b>III.1.1</b> Students (A) will use source sheets (C) to demonstrate an understanding of each source's main ideas. (B)	Research summary shows a lack of understanding of the authors' main points; <i>and/or</i> Research summary shows a lack of direct connection between sources and the student's thesis.	Source evaluations accurately summarize each author's main points; <i>but</i> Source evaluations do not demonstrate a clear connection between the authors' purposes and the student's thesis; <i>and/or</i> Sources selected are on the same general topic but lack a cohesive point.	Sources are all related to the student's thesis. Source descriptions show a good understanding of the authors' purposes.	Sources comprise a cohesive body of research which show a clear theme running from one source to another; <i>and</i> Descriptions show a clear understanding of what each author was trying to accomplish; <i>and</i> Sources do not make identical points but support and challenge the student's thesis with various congruent arguments.
<b>CITATION</b> <b>II.5.3 V.3.1</b> On research project source sheets (C) students (A) will use an assigned citation format to create complete and accurate citations for sources used. (B)	Citations give the reader insufficient information to locate the original source.	Citations show substantial errors in format but would not prevent the reader from locating the source.	Citations have only minor errors in form, such as misplaced punctuation, that would not prevent the reader from locating the source.	Citations are consistently accurate in both content and form for sources in various formats.
<div style="border: 1px solid black; padding: 5px;"> <p style="text-align: center;"><b>KEY</b></p> <p>Numbers for each objective refer to ACRL Standards            Objectives are developed using the A-B-C-D method:            A=Audience being assessed            B=Behavior to be performed.            C=Conditions under which behavior will be performed            D=Degree to which behavior will be performed, defined by the rubric</p> </div>				

*Figure 4.13. Kobritz's Information Literacy Rubric*

and the instructor assessed early and final drafts of students' research papers as well as a reflection paper students wrote after completing the research paper.<sup>1237</sup>

Using the same rubric across multiple performances, Warmkessel demonstrated that all students made progress through the drafts of their research essays and "most exhibited through their reflective essays an understanding of the need to find current and authoritative sources to be integrated into their essays."<sup>1238</sup>

Warmkessel notes that rubric assessment revealed students' misunderstandings of which she and the instructor had previously been unaware.<sup>1239</sup> She concludes

<sup>1237</sup> Warmkessel, "Assessing Abilities of Freshmen." 251.

<sup>1238</sup> Warmkessel, "Assessing Abilities of Freshmen." 251.

<sup>1239</sup> Warmkessel, "Assessing Abilities of Freshmen." 252-253.

that rubric assessment, while time-consuming, was an effective way to assess students' information literacy skills.<sup>1240</sup>

In a different 2003 publication, the Bay Area Community Colleges released information about their Information Literacy Competency Project.<sup>1241</sup> This project includes the development and field-testing<sup>1242</sup> of a two-part assessment of information literacy skills. This assessment is used as a "challenge-out exam, or credit by examination, to provide students with a way to satisfy an information competency [graduation] requirement" for an associate's degree.<sup>1243</sup> The authors state clearly that this assessment is not sufficiently field-tested to establish validity or reliability.<sup>1244</sup>

The first part of the Bay Area Community Colleges Information Competency Assessment Project is a 47-item test and, like many other standardized test approaches to information literacy assessment, it is primarily multiple-choice and matching items. Of the 47 items, seven are short answer, constructed-response questions. To score these items, librarians used a rubric (see Figure 4.14).

The second half of the Bay Area Community Colleges Information Competency Assessment Project, however, is comprised of 12 performance-

---

<sup>1240</sup> Warmkessel, "Assessing Abilities of Freshmen." 252.

<sup>1241</sup> Smalley, Bay Area Community Colleges.

<sup>1242</sup> Andy Kivel, "Institutionalizing a Graduation Requirement," Assessing Student Learning Outcomes for Information Literacy Instruction in Academic Institutions, ed. Elizabeth Fuseler Avery (Chicago: Association of College and Research Libraries, 2003).

<sup>1243</sup> Smalley, Bay Area Community Colleges.

<sup>1244</sup> Smalley, Bay Area Community Colleges.

Test Item	Scoring: Answers and points
39. What is an example of a primary source of information?	<p><b>2 pts. for acceptable answer; 0 for unacceptable</b>            Examples of <b>acceptable</b> answers: diary, letter, speech, autobiography; eyewitness account; research report; scholarly journal article reporting original research; novel; interview            Examples of <b>unacceptable</b>: secondary sources of information; scholarly article; <b>no partial score</b></p>
41. What disciplinary action might happen to community college students who violate college policies on plagiarism?	<p><b>2 pts. for acceptable answer; 0 for unacceptable</b>            Examples of <b>acceptable</b>: fail the course; fail the assignment; meeting with instructor/chair/dean; suspension from college, etc.            Examples of <b>unacceptable</b>: probably nothing, re-do the assignment. <b>no partial score</b></p>
42. You need to read the novel _____, but the book is checked out from your college library. You don't have the money to buy it from a bookstore, so how can you get a library copy of the book?	<p><b>2 pts. for acceptable answers; 0 for unacceptable</b>            Examples of <b>acceptable</b>: ILL/document delivery; place a hold on the book; go to another library; check another library's opac; ask a friend/instructor if he/she has this novel; ask a librarian for help; etc.            Examples of <b>unacceptable</b>: borrow the money to buy a copy; search the Web. <b>no partial score</b></p>
43. Popular periodicals like <i>Newsweek</i> usually contain many advertisements. What are 2 other characteristics of popular periodicals?	<p><b>1 pt. for each acceptable feature for total of 2 pts. possible scores; 0, 1, or 2 pts.</b>            Examples of <b>acceptable</b> answers: appeal to a more popular audience; have more pictures/illustrations; glossy paper; shorter articles; usually no bibliographies or footnotes; language/content less specialized and easier to understand; usually no abstracts; not refereed; cover a wider range of topics/issues in one issue; often more frequent publication; usually not written by experts/researchers; etc.            Examples of <b>unacceptable</b>: written for the average person/layperson; more fun; cover topics I'm interested in, etc.</p>
44. Scholarly periodicals like <i>American Sociological Review</i> usually contain very few advertisements. What are 2 other characteristics of scholarly periodicals?	<p><b>1 pt. for each acceptable feature for total of 2 pts. possible scores; 0, 1, or 2 pts.</b>            Examples of <b>acceptable</b> answers: appeal to a more scholarly/academic audience; lengthy articles; often have footnotes or bibliographies; written by experts/professionals; usually include abstracts; may contain research findings; may be less frequently published; language/content more specialized and technical/research-oriented; usually fewer pictures/illustrations; refereed, etc.            Examples of <b>unacceptable</b>: written for experts/scholars/researchers; hard to understand; boring; etc.</p>

Test Item	Scoring: Answers and points
<p>45. Your research question is “Do elementary school students receive adequate science instruction in California public schools?” How could you broaden this question? Write an example of a broader research question.</p>	<p>Criteria to apply:</p> <p><b>Criterion 1</b> – broadened question is within or relates to the research question’s subject.</p> <p><b>AND</b></p> <p><b>Criterion 2</b> – Question is broadened by specifying broader geographic area, broader student group; etc.</p> <p><b>AND</b></p> <p><b>Criterion 3</b> - Research question contains a subject and is phrased as a recognizable type of question, such as a question of fact; a question that posits a comparison/contrast or a probing/investigative question.</p> <p><b>All 3 criteria = 2 points</b>  <b>2 out of 3 criteria = 1</b>  <b>1 out of 3 criteria = 0</b></p> <p><b>possible scores 0, 1, 2</b></p> <p>Examples <b>acceptable</b>:  --Do K-12 students receive adequate science instruction in public schools?</p> <p>Examples of <b>unacceptable</b>:  --Do students receive good instruction ?  --Do elementary school teachers receive adequate science instruction training? (doesn't relate closely enough to the research question)</p>
<p>46. You need to make a presentation to the City Council meeting against the proposed regulation of banning dogs from city parks. What are <b>two</b> approaches you could take to find information supporting your point of view? You have already used Web search tools and databases to find some information. You still need more information so you want to use <b>other methods of inquiry</b>.</p>	<p><b>1 point for each acceptable answer; 0 for unacceptable</b></p> <p><b>Possible scores: 0, 1, 2 points</b></p> <p>Examples of <b>acceptable</b>: conduct a survey; interview neighbors/citizens; obtain opinions of dog-walking groups or other organizations supportive of dogs in the parks; telephone or e-mail other nearby cities to find out their policies; find newspaper articles or editorials in support; poll people in dog lovers chat rooms; etc.</p> <p>Examples of <b>unacceptable</b>: do a Web search; find articles on the Web or in databases; use Google to find good supporting information; etc.</p>
<p>47. To gain access to your college library’s online databases from off-campus you would:</p>	<p><b>2 points for acceptable; 1 point for partially acceptable; 0 points for unacceptable</b></p> <p><b>possible scores: 0, 1, 2</b></p> <p>Examples of <b>acceptable</b>: need to have an Internet connection and then connect to the library’s Website or other URL where there are links to the online databases. I would then type in a password or user ID or barcode, etc. (2 steps included connecting to library Web site and using password.)</p> <p>Example of <b>partially acceptable</b>: type in a password or user ID or barcode;</p> <p>Example of <b>unacceptable</b>: connect to the Library’s Website to access them; use my home computer to access the Library’s databases (zero points if no mention of needing a password or ID number or barcode)</p>

*Figure 4.14. Bay Area Community Colleges’ Information Literacy Rubric #1*

based exercises that are evaluated using a complex rubric (see Figure 4.15).

Students are asked to narrow a broad topic, select key concepts from

Test item	Criteria to apply	What constitutes competent	Notes on scoring	Score, and weighted score
<b>EXERCISE I – B.1.a Narrowed topic</b>				
B.1.a Your instructor has given you the broad subject <b>civil rights in America</b> and an assignment to write a 3-5 page research paper on some aspect of this subject. Narrow this subject to a manageable topic for the assignment.	<p>B.1.a Criterion 1 – Topic is narrowed by specifying time frame, or persons, or organization or group, or location, or event or incident, or some combination of these, or other similar, appropriate limiter(s) is (are) applied. <b>AND</b></p> <p>Criterion 2 – Narrowed topic is within subject assigned. <b>AND</b></p> <p>Criterion 3 – Narrowed topic is appropriate to a 3-5 page research paper.</p>	<p>B.1.a At least one of the narrowing techniques has been applied to the subject. <b>AND</b> The other 2 criteria are met.</p>	<p>B.1.a Broad topic is civil rights in America Examples of appropriately narrowed topics ✓ <i>Status of Arab American Civil Rights After 9/11</i> [narrowed by time frame and group] ✓ <i>Effects of the UC system prohibiting race-based admission criteria</i> [narrowed by organization and event] Example of a topic not properly narrowed ✓ <i>The civil rights movement in America, equality then and now</i> [no time period, incident, or group used to focus] Example of topic that is <i>not</i> within subject assigned ✓ <i>The two sides fighting the Civil War</i>  Examples of appropriate topics for a 3-5 page paper ✓ <i>Civil rights of American-Japanese during World War II</i> ✓ <i>Black American civil rights from 1960-1966</i> Examples of topics that are too broad ✓ <i>The rights of Americans</i> ✓ <i>Black Americans' civil rights</i></p>	<p>B.1.a Criterion 1=1 Criterion 2=1 Criterion 3=1  3 out of 3 = 1 0, 1, or 2 out of 3 = 0  Possible scores 1, 0  <b>Weighted score</b> 1 = 6 0 = 0</p>
<b>B.1.b Steps taken to develop narrowed topic</b>				
B.1.b In two or three sentences, describe in detail what you did to develop your narrowed topic.	<p>B.1.b Criterion 1 – The student describes the process used to consult additional resources, such as background information sources, to develop a narrowed topic. <b>AND/OR</b></p> <p>(continued on next page)</p>	<p>B.1.b What the student writes indicates that the process used meets Criterion 1 <b>AND/OR</b> Criterion 2</p>	<p>B.1.b The response should be in sentences, or, at the very least, easily understandable. Example of an appropriate response ✓ <i>I went to the Encyclopedia Americana and looked up "Civil Rights and Liberties." The subsections of the big article were laid out in such a way that I could identify a narrower topic that interested me.</i> Example of an inadequate response: ✓ <i>First I brainstormed possible specific ideas. Then I narrowed the ideas down.</i> ✓ <i>I went to Alice and typed civil rights in America and got a lot of results.</i> (continued on next page)</p>	<p>B.1.b Criterion 1 = 1 Criterion 2 = 1  1 or 2 = 1 0 = 0  Possible scores 1, 0  <b>Weighted score</b> 1 = 6 0 = 0</p>
<b>Test item</b>	<b>Criteria to apply</b>	<b>What constitutes competent</b>	<b>Notes on scoring</b>	<b>Score, and weighted score</b>
	<p>Criterion 2 – The student describes the process used to implement his/her own prior knowledge to develop a narrowed topic.</p>		<p>Example of an appropriate response ✓ <i>I tried to think of a specific incident in the civil rights movement. I remembered Rosa Parks and decided to focus on her famous action in the bus and its consequences. What were the consequences of her action?</i></p>	
<b>B.2 Posing the research question</b>				
B.2 Next, take your narrowed topic and pose it as a research question that you could adequately address in this 3-5 page writing assignment.	<p>B.2 Criterion 1 – The student composes a research question that contains a subject. <b>AND</b> Criterion 2 – The question is a recognizable type of question (such as a question of fact; a posited hypothesis; some topic for comparing/contrasting; or it is a probing or investigative question; etc.) <b>AND</b></p> <p>Criterion 3 – The research question is within the narrowed topic <b>AND</b></p> <p>Criterion 4 – The research question is appropriate to the assignment</p>	<p>B.2 The research question meets ALL of the 4 criteria.</p>	<p>B.2 Example of a research question that does not contain a subject ✓ <i>What are the effects?</i>  Examples of appropriate responses ✓ <i>What arguments did the lawyers make in the 1954 Brown v. Board of Education Supreme Court case?</i> [question of fact] ✓ <i>How has the tragedy of September 11 affected the civil rights of Arabs and Muslims living in America?</i> [investigative question]  Example of inappropriate response ✓ <i>Black movement in respect to civil rights.</i> [not a research question]  Look back at the narrowed topic – is the research question within that topic?  Keeping in mind the 3-5 page research paper assignment, is the research question appropriate?</p>	<p>B.2 Criterion 1 = 1 Criterion 2 = 1 Criterion 3 = 1 Criterion 4 = 1  4 out of 4 = 1 0, 1, 2, or 3 out of 4 = 0  Possible scores 1, 0  <b>Weighted score</b> 1 = 6 0 = 0</p>
<b>EXERCISE II - B.3.a.1, b.1.c.1 Three key concepts</b>				
B.3.a.1, b.1.c.1 You've been given the assignment to write a 3-5 page research paper on the following question: <b>Should colleges be allowed to restrict student speech?</b> Write in the key concepts represented by the research question.	<p>B.3.a.1, b.1.c.1 The student identifies 2 or 3 distinct main concepts represented by the research question.</p>	<p>B.3.a.1, b.1.c.1 The student identifies 2 or 3 distinct main concepts which are 1) colleges or higher education; 2) speech; 3) restriction or censorship.</p>	<p>B.3.a.1, b.1.c.1 Appropriate responses <i>Key concept 1: colleges</i> <i>Key concept 2: speech</i> <i>Key concept 3: censorship</i></p>	<p>B.3.a.1, b.1.c.1 Each distinct key concept = 1 2 or 3 out of 3 = 1 0 or 1 out of 3 = 0 Possible scores 1, 0 <b>Weighted score</b> 1 = 6 0 = 0</p>

Test item	Criteria to apply	What constitutes competent	Notes on scoring	Score, and weighted score
<b>B.3.a.2-3,b.2-3,c.2-3 Synonyms or alternate terms</b>				
B.3.a.2-3,b.2-3,c.2-3 Then write in synonyms or alternate terms for each concept that you would use in searching for information on this question.	B.3.a.2-3,b.2-3,c.2-3 The student identifies a total of 4 (out of a possible 6) appropriate alternate terms or synonyms for the key concepts.	B.3.a.2-3,b.2-3,c.2-3 The response meets the criterion.	B.3.a.2-3,b.2-3,c.2-3 Examples of appropriate responses <i>Key concept 1 synonym 1: universities</i> <i>Key concept 1 synonym 2: higher education</i> <i>Key concept 2 synonym 1: free expression</i> <i>Key concept 2 synonym 2: 1st amendment rights</i> <i>Key concept 3 synonym 1: restrain</i> <i>Key concept 3 synonym 2: restrict</i>	B.3.a.2-3,b.2-3,c.2-3 Each appropriate alternate term or synonym = 1  4, 5 or 6 out of 6 = 1 0, 1, 2, or 3 out of 6 = 0  Possible scores 1, 0  Weighted score 1 = 6 0 = 0
<b>B.4* Style guide</b>				
B.4* What citation format will you be using (e.g., MLA, APA, CBE)?		B.4* Style guide is specified.	B.4* Style guide is specified in such a way that it is obvious which format is being utilized. Edition statement not required.	B.4* Style guide is appropriately specified = 1 Not specified, or not specified appropriately = 0  Possible scores 1, 0  Weighted score 1 = 1 0 = 0

*Figure 4.15. Bay Area Community Colleges' Information Literacy Rubric #2*

a supplied research question, choose relevant sources for a research question, cite sources, and evaluate websites.<sup>1245</sup> Students enter their responses to these tasks into a web form, and a rubric is used to assess their responses. Smalley reports that the conceptualization and development of the scoring rubric was the most difficult part of developing the second half of the assessment project.<sup>1246</sup> One area of difficulty was describing the attributes of acceptable and unacceptable performances.<sup>1247</sup> The original rubric was revised after the first field test using actual student responses, and the three-level scoring concept was eliminated, since the purpose of the assessment was to determine if students could demonstrate an acceptable level of performance to meet the graduation requirement.<sup>1248</sup> The finished rubric includes scoring criteria, guidelines, and

<sup>1245</sup> Smalley, Bay Area Community Colleges.

<sup>1246</sup> Smalley, Bay Area Community Colleges.

<sup>1247</sup> Smalley, Bay Area Community Colleges.

<sup>1248</sup> Smalley, Bay Area Community Colleges.

examples of appropriate and inappropriate responses.<sup>1249</sup> Unfortunately, the current version of the rubric still is difficult to read and so task-specific that it functions more as an answer key than a rubric.

Smalley recommends that three different scorers assess student responses to these items.<sup>1250</sup> She also recommends training scorers to improve interrater reliability.<sup>1251</sup> While Smalley does not go into further detail on this point, she does suggest that raters read and discuss the rubric, then practice scoring, and discuss their interpretations of a few exams.<sup>1252</sup>

---

<sup>1249</sup> Smalley, Bay Area Community Colleges.

<sup>1250</sup> Smalley, Bay Area Community Colleges.

<sup>1251</sup> Smalley, Bay Area Community Colleges.

<sup>1252</sup> Smalley, Bay Area Community Colleges.



## **CHAPTER 5**

### **METHODOLOGY**

This study employed a survey design methodology. The data for the study came from student responses to open-ended questions embedded in an online information literacy tutorial. This textual data was translated into quantitative terms through the use of a rubric. Using a rubric, raters coded student answers into pre-set categories, and these categories were assigned point values. The point values assigned to student responses were subjected to quantitative analysis in order to describe student performance, test for interrater reliability, and explore the validity of the rubric. According to Lincoln, this approach is called “discovery phase” or preliminary experimental design, and it is commonly employed in the development of new rubrics.<sup>1253</sup>

#### **Study Participants**

This study includes two large groups of participants: (1) the students who provided artifacts of student learning by responding to writing prompts in an online information literacy tutorial, and (2) raters who scored the students’ responses.

---

<sup>1253</sup> Yvonna Lincoln. "Authentic Assessment and Research Methodology." E-mail to Megan Oakleaf. 2005.

### ***Providers of Artifacts of Student Learning***

One of the major challenges faced by academic librarians seeking to teach information literacy skills is the diversity of skills found in first-year college students. In order to teach college-level information literacy skills, academic librarians must first establish baseline knowledge on which to build more complex skills. At North Carolina State University (NCSU), the LOBO tutorial seeks to supply incoming students with basic information literacy skills. At other institutions, librarians use one-shot workshops, for-credit classes, and other approaches to address the needs of first-year students. Whatever the approach to establishing these basic skills, librarians often question whether or not the measures taken are in fact meeting the goal of establishing baseline information literacy skills. Because so many academic librarians struggle with this challenge, this study uses artifacts of student learning produced by an approach to this problem.

In the fall of 2004, NCSU admitted 3,800 first-year students, and 97% of them enrolled in English 101, a first-year composition course, during either fall or spring semester of the 2004-2005 academic year. As a part of their English 101 coursework, all students completed the information literacy online tutorial called LOBO. As students progressed through the LOBO tutorial, they answered open-ended questions designed to capture evidence of student learning. These answers were stored in a database on library servers. As a result, the LOBO answer database provided a wealth of information for the assessment of library instruction.

During spring semester 2005, 800 students responded to the LOBO question that was examined in this study. Of the 800 student responses, a small number of responses could not be scored due to blank entries or lack of adherence to directions. From the remaining responses, a random sample was selected for analysis using the rubric. No special student characteristics were sought; it was expected that by using random selection, a reasonable variation in student gender, race or ethnicity, and academic achievement would occur.

### ***Raters***

Twenty-five raters participated in this study. The raters were evenly divided into five groups: NCSU librarians, NCSU English 101 instructors, NCSU English 101 students, instruction librarians from other ARL libraries, and reference librarians who had some instruction responsibilities from other ARL libraries. These rater groups were chosen for a variety of reasons.

NCSU librarians were selected as a rater group for this study because they are responsible for the instruction of information literacy skills at NCSU, they are interested in ways to assess student learning, and they have the most to gain from participating in the study. The specific librarians chosen as raters were selected in an attempt to represent a cross section of reference and instruction librarians in gender and race. At the NCSU Libraries, 56 percent of librarians were female in 2005.<sup>1254</sup> According to 2000 American Library Association (ALA)

---

<sup>1254</sup>Laura Blessing. "Gender and Race in the NCSU Libraries." E-mail to Megan Oakleaf. 10 August 2005.

statistics, 82.3 percent of all credentialed librarians were female.<sup>1255</sup> In the study, there were four female librarians and one male librarian, so the gender composition of this rater group resembles the national percentages of gender more closely than the NCSU percentages. In 2000, 89.9 percent of credentialed librarians were white,<sup>1256</sup> and at NCSU, 85.2 percent of librarians self-identified as white in 2005.<sup>1257</sup> In the study, there were four Caucasian librarians and one African-American librarian so the racial background of NCSU librarians involved in the study approximates the backgrounds of librarians both nationally and locally.

The NCSU librarian rater group also represented a variety of areas of expertise. One librarian was a veteran of twenty-two years, a subject specialist in the humanities, and a former coordinator of the library instruction program. At the time of the study, this librarian's responsibilities included delivering instructional presentations to classes, creating web pages to facilitate student discovery of library resources, and developing course assignments in conjunction with instructors. A second NCSU librarian also served as the coordinator of instruction during her eight years as a professional librarian, but she was the assistant head of the reference department at the time of the study. In this role, she trained staff to work at the reference desk and assessed the content learned in these training sessions. The third librarian was a subject specialist librarian of

---

<sup>1255</sup> Tracie Hall. "Question About Minority Makeup of Library Profession." E-mail to Megan Oakleaf. 18 August 2005.

<sup>1256</sup> Hall. "Question About Minority Makeup of Library Profession."

<sup>1257</sup> Blessing. "Gender and Race in the NCSU Libraries."

six years who was tasked with teaching and assessing the learning of undergraduate engineering students. The fourth was a subject specialist librarian serving the college of education. This librarian taught library instruction classes for six years, but did not attempt to formally assess that instruction. Finally, the fifth librarian served as a reference librarian for distance learning students for six years and also held a management position in the reference department at the time of the study. This librarian supervised other librarians who had instruction responsibilities. Two of these librarians participated in the creation of the LOBO tutorial. The NCSU librarians participating in this study were recruited through informal conversations about the study.

English 101 instructors were selected as another rater group because they were co-creators of the LOBO tutorial and had a vested interest in the tutorial's effectiveness. Longtime stakeholders in information literacy skills instruction, they had sought methods to assess their students' use of LOBO in the past. The five instructors participating in this study were recruited using an email message to the English 101 instructor listserv and announcements at First-Year Writing Program meetings. The individual instructors were chosen in an attempt to reflect the composition of all ENG 101 instructors. In 2005, 94 percent of ENG 101 instructors were white, and 60 percent were female.<sup>1258</sup> In the study, all ENG 101 instructors were white and four were female. The instructors in this group had varying levels of experience teaching the course and using LOBO. One ENG 101 instructor was a graduate student, three were ranked as Lecturer

---

<sup>1258</sup> Ann M. Penrose. "Race and Gender in the First Year Writing Program." E-mail to Megan Oakleaf. 10 August 2005.

I, and one was a Lecturer II. Their experience teaching ENG 101 ranged from 1 semester to 20 semesters. None contributed directly to the content of the LOBO tutorial, but all five used LOBO to teach their ENG 101 courses.

Five students were selected as raters to ensure student input into the rubric assessment process. The students who participated in this study were enrolled in ENG 101 during the previous semester. The students who served as raters in this study volunteered to participate in response to announcements made by their former ENG 101 instructors. Because 56 percent of the first-year population for the 2004-2005 academic year at NCSU was male, three of the five student raters were male. Although 81 percent of the first-year students in 2004-2005 were white and 10.7 percent were African-American, in this study three students were white and two were African-American. Each student represented a different college of the university. Their majors include accounting, communications, computer science, and mechanical engineering; one student's major was undecided. Four students self-reported that they earned a B in ENG 101; the remaining student earned an A.

Finally, ten librarians at other campuses were included in two additional rater groups to explore the performance of raters outside the NCSU campus context. All ten librarians worked at five ARL library systems. Each of the five library systems was represented by an instruction librarian and a reference librarian who had some instruction responsibilities. All external (non-NCSU) raters responded to calls for participants circulated on the Information Literacy Instruction listserv.

The five external instruction librarians who volunteered and were selected for this study self-identified as white females. All were responsible for delivering library instruction, and their job titles included Coordinator for Instruction and Outreach, Instruction Department Head, and Coordinator for Information Literacy Services and Instruction. One had just completed her first year as a librarian, three were librarians for 11-14 years, and one served as a librarian for 23 years. Three helped create or maintain information literacy tutorials. Two were not responsible for assessment of learning, and the remaining three external instruction librarians cited satisfaction surveys and pre- and post-tests as their main assessment tools. None had experience with rubrics.

The five external reference librarians were all female. Four self-identified as white, one as Asian. Three had contributed to the creation or maintenance of an online information literacy tutorial, and all five had library instruction responsibilities. Four of the five were not responsible for the assessment of instruction. Two of the five had used rubrics on a limited basis in the past.

### **The Rubric**

The rubric used in this study sought to capture students' ability to use authority as a criterion for evaluating a website. The student outcomes assessed by the rubric were based on an adaptation of the Information Literacy Competency Standards for Higher Education and the Objectives for Information Literacy Instruction: A Model Statement for Academic Librarians. The first adapted outcome is: "The student will articulate established evaluation criteria."

This outcome is based on the ACRL standard 3.2, which states, “The information literate student articulates and applies initial criteria for evaluating both the information and its sources” and standard 3.2.a, which reads, “The information literate student examines and compares information from various sources in order to evaluate reliability, validity, accuracy, authority, timeliness, and point of view or bias.”<sup>1259</sup> The second adapted outcome is: “The student will apply criteria to analyze information, including authority, to information and its source.” This outcome was adapted from ACRL standard 3.2.c, which is: “The information literate student applies evaluative criteria to information and its source.” A third adapted outcome was derived from ACRL standard 3.2.a. The outcome reads, “The student will investigate an author’s qualifications and reputation.” The fourth adapted outcome states, “The student will evaluate sources for use.” This outcome was derived from ACRL standard 3.4.g which states that students should (1) describe “why not all information sources are appropriate for all purposes,” (2) distinguish “among various information sources in terms of established evaluation criteria,” and (3) apply “established evaluation criteria to decide which information sources are most appropriate.”<sup>1260</sup> The last adapted outcome states, “The student will indicate whether or not a specific, individual source is appropriate for the purpose at hand, based on established evaluation criteria, and provide a rationale for that decision.” This outcome was based on

---

<sup>1259</sup> Association of College and Research Libraries, Objectives for Information Literacy Instruction: A Model Statement for Academic Librarians, 2001, 25 March 2005  
<<http://www.ala.org/ala/acrl/acrlstandards/objectivesinformation.htm>>.

<sup>1260</sup> Association of College and Research Libraries, Objectives for Information Literacy Instruction.



ACRL standards 3.2.c and 3.4.g described above. (For a full set of LOBO outcomes, see Appendix A.) The rubric included four criteria and three levels of performance. The criteria listed in the rubric were: “Articulates Criteria”, “Cites Indicators of Criteria”, “Links Indicators to Examples from Source”, and “Judges Whether or Not to Use Source”. The rubric also described student behavior at three levels: Beginning, Developing, and Exemplary. The rubric was revised numerous times based on feedback from NCSU assessment professionals and pilot testing. (For the version of the rubric used in the study, see Appendix B.)

### **The Pilot Test**

A preliminary test of the rubric was conducted during spring semester 2004. The purpose of this pilot test was to make improvements both to the rubric and to the open-ended questions that form the writing prompt for students. By improving the prompt, the researcher sought to increase the level of detail in the student responses in order to ensure an adequate basis for the present study. By clarifying and fine-tuning the rubric, the researcher also sought to facilitate the use of the rubric by other raters.

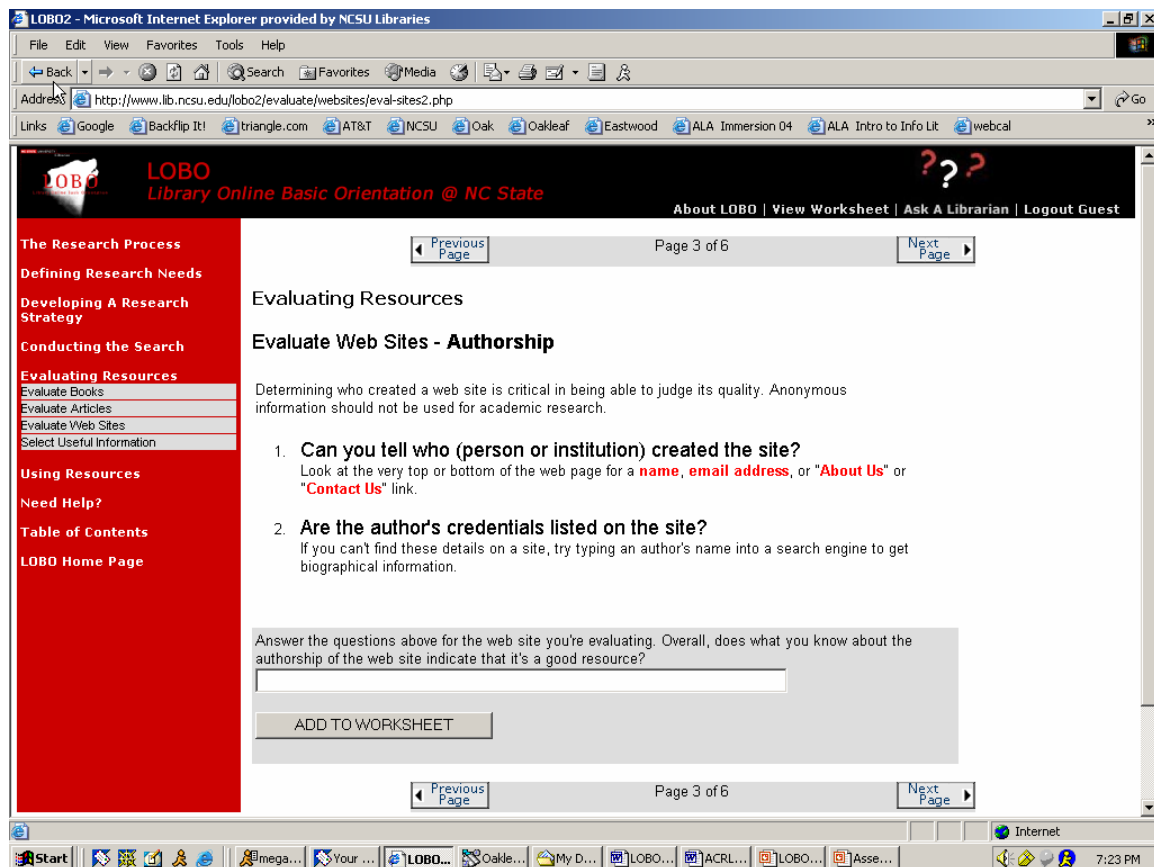
In the pilot study, one entry-level librarian used the rubric to score fifty student responses to the LOBO tutorial prompt that elicits information about students’ abilities to use authority as a criterion for evaluating a website. This prompt required students to respond to a series of questions about the authority of a website they were considering using for an academic paper or project. The fifty student responses were selected randomly from the LOBO answer database

and separated from any personally identifying information. The librarian scored all fifty responses. He found that a majority of students scored an “Exemplary” for the first criterion on the rubric. This indicates that students were able to address the authority of a website (88%). Most students also scored an Exemplary on the second criterion of the rubric, demonstrating that they were able to refer to indicators of authority (90%). However, only less than a third (32%) of students scored an Exemplary on the third rubric criterion and could give specific examples of authority indicators from the site they were evaluating. Fewer than half (44%) scored an Exemplary on the fourth criterion and could provide a rationale for accepting or rejecting the website for use in their assignment based on their assessment of the site’s authority. The results of the pilot test were used as anticipated—to improve the rubric, the content of the tutorial, and the open-ended questions that form the writing prompt (See Figures 5.1 and 5.2). After the pilot test, a version of the rubric was created that used more student-accessible language. In order to make future assessments more “transparent” to students, the rubric was included in the tutorial under a link labeled “How might an instructor score your answer?”

### **Study Procedure**

The procedures followed in this study can be divided into four parts. First, the artifacts of student learning and study materials were prepared. Then, the three internal rater groups met in person for a 6-hour training and scoring


session. Third, the two external rater groups received, scored, and returned their study materials. Finally, the score sheets were prepared for statistical analysis.



*Figure 5.1. LOBO Tutorial Before Pilot Testing  
North Carolina State University Libraries*

### **Preparation of Study Materials**

Student responses to the LOBO tutorial writing prompt were prepared using a multi-step process. First, all student responses were retrieved from the LOBO answer database and separated from personally identifying information. After null and unscorable responses were removed, the remaining 800 responses were numbered consecutively. Using a random number table, 75 student responses were selected for the study.



**LOBO**  
Library Online Basic Orientation @ NC State

???

About LOBO | View Worksheet | Ask A Librarian | Logout Guest

Previous Page

Page 2 of 6

Next Page

**The Research Process**

Defining Research Needs

Developing A Research Strategy

Conducting the Search

**Evaluating Resources**

Evaluate Books

Evaluate Articles

Evaluate Web Sites

Select Useful Information

Using Resources

Need Help?

Table of Contents

LOBO Home Page

**Evaluating Resources**

**Evaluate Web Sites - Authority**

The URL (web address) and author information for a web site reveal a lot about site reliability. Determining who created a web site is critical in being able to judge its quality. Generally, anonymous information should not be used for academic research.

Consider the following questions when you're evaluating the authority of a web site:

- What type of domain does the site come from?**  
Government sites use **.gov** and **.mil** domains. Educational sites use the **.edu** domain. Non-profit organizations use **.org** and business sites use **.com**. Generally, **.gov** and **.edu** sites are considered more trustworthy than **.org** and **.com** sites.
- Who "published" the site?**  
The name between **http://** and the first **/** usually indicates what organization owns the server the web site is housed on. Learning about the organization that hosts a site can give you important information about the site's credibility.  
**http://www.wired.com/news/news/**
- Is it a personal web site?**  
Look for the names of companies that sell web space to individuals, like AOL or GeoCities. Also look for a tilde (~). Tildes are often used to signify a personal web site. Personal sites are considered less reliable than sites supported by organizations.
- Can you tell who (person or institution) created the site?**  
Look at the very top or bottom of the web page for a **name**, **email address**, or **"About Us"** or **"Contact Us"** link.
- Are the author's credentials listed on the site?**  
If you can't find these details on a site, try typing an author's name into a search engine like Google to get biographical information.

**Respond to the following prompts in the space below, using complete sentences:**

- Identify the "domain type" of the site you're evaluating and explain why that is acceptable or unacceptable for your needs.
- Identify the "publisher" or host of the site and tell what you know (or can find out) about it.
- State whether or not the site is a personal site and explain why that is acceptable or unacceptable for your needs.
- State who (name the person or institution) created the site and tell what you know (or can find out) about the creator.
- Look for the author's credentials on the site. List his/her credentials and draw conclusions based on those credentials. If there are no credentials listed, tell what conclusions you can draw from their absence.
- Using what you know about the **AUTHORITY** of this web site, explain why it is or is not appropriate to use for your paper/project.

ADD TO WORKSHEET

How might an instructor score your answer?

Previous Page

Page 2 of 6

Next Page

Ask a Librarian | Copyright | Disclaimer

Last Modified: 01/07/05 12:06pm  
Questions/Comments to [LibWebTeam](#)  
URL: <http://www.lib.ncsu.edu/lobo2/evaluate/websites/eval-sites1.php>

Figure 5.2. LOBO Tutorial After Pilot Testing  
North Carolina State University Libraries

Each of the 75 responses was placed on a score sheet (see sample score sheet in Appendix C). Score sheets included the student response, the scoring rubric, and three code numbers: the number of the response, the position of the response among the 75 to be scored, and the rater's number.

Next, the researcher scored each of the 75 responses three times using the study rubric. Afterwards, the researcher reviewed the scores assigned to each student response and reconciled any divergent scores. This process was necessary to ensure the validity of the “gold standard” approach explained later in this chapter.

After each response was assigned a score, the researcher sorted the student responses into groups according to its score so that all responses with a score of “8” were together, etc. Then, the researcher created three large groups of 25 student responses, making sure to include an equal number of high, medium, and low scoring responses in each group of 25. Finally, within each group of 25 responses, individual responses were arranged in their original random order. This process resulted in three separate groups of student responses with each group having an even number of high, medium, and low scoring responses included. The three separate groups were numbered 1-25, 26-50, and 51-75 and distributed to raters in this order. This approach enabled the researcher to later examine the reliability with which raters scored the first third, middle third, and last third of the student responses.

Twenty-five student responses were also chosen from the LOBO database for other purposes. Fifteen responses were selected as model or

“anchor” papers to be used in the training session for internal raters. Ten additional responses were selected as a “practice” test taken by all raters before scoring the 75 randomly selected student responses. The fifteen anchor responses and ten practice test responses were not chosen randomly. Rather, they were selected because they represented the wide range of student responses included in the study sample. All these responses were placed on score sheets identical to those used in the study.

Because the research design employed in this study was based on the realities of assessment in academic libraries, the preparation of study materials for internal and external raters differed. In academic libraries, assessments are either created on campus where training is available, or assessments are imported from a separate institution and only written materials are available for consultation. In this study, the internal raters participated in a training session, and the external raters were provided with a substantial amount of background material, directions, and examples to familiarize them with the campus culture and study context.

Materials prepared for the internal rater training included a meeting agenda, consent forms, rater data forms, poster-sized versions of the rubric, a Power Point presentation (see Appendix D), copies of the ACRL Information Literacy Competency Standards for Higher Education, copies of LOBO Information Literacy Skills Objectives and Outcomes, screenshots of LOBO, and open-ended comment sheets to be completed by raters at the close of the scoring session.

Materials prepared for the external rater mailing included several handouts: an inventory of materials, the context of the study, consent forms, rater data forms, directions for scoring the 10 practice student responses, directions for scoring the 75 study responses, open-ended comment sheets to be completed by raters at the close of their study participation, return mail checklists, and postage paid return mail envelopes. (For selected external rater materials, see Appendix E.)

### ***The Internal Rater Experience***

The internal rater section of this study was conducted in one session during which the researcher met with 5 NCSU librarians, 5 NCSU English 101 instructors, and 5 NCSU English 101 students in one 6-hour period. One graduate assistant also supported the session by documenting the receipt of raters' paperwork and coordinating refreshments.

As raters entered the testing session, the researcher and graduate assistant welcomed them. Raters were instructed to make nametags and locate their seat. Raters were divided into five small groups, each seated at a different table. Groups consisted of one librarian, one ENG 101 instructor, and one ENG 101 student to bring out diversity of opinion during the training session and to prevent rater groups (librarians, instructors, and students) from engaging in "group think". Meeting agendas, consent forms, rater data forms, invoice paperwork (for raters' stipends), copies of the ACRL Information Literacy Competency Standards for Higher Education, copies of LOBO Information

Literacy Skills Objectives and Outcomes, and screenshots of LOBO were available at each group's table. Once all raters were in attendance, the researcher led them through the completion of their consent forms, rater data forms, and invoice paperwork. After the necessary paperwork was collected, all attendees introduced themselves briefly and were formally welcomed.

Next, the researcher explained the purpose of the study, defined information literacy, and described the need for tools to assess information literacy skills. Because most of the raters had little prior experience with rubrics, the researcher introduced rubrics by providing a definition, describing the component parts (criteria and performance levels), and providing brief examples of rubrics. She also reviewed the relevant sections of LOBO including the content that is conveyed through the tutorial and the open-ended questions that serve as a writing prompt for students. The researcher described the origins of the outcomes espoused by the study rubric and explained the relationship between the outcomes and the rubric criteria and performance levels. Finally, she showed raters the student version of the study rubric that is viewable online in LOBO (see Appendix F). Raters acknowledged that the student version of the rubric, written in more accessible language, included the same content as the study rubric. The researcher closed this section of the training by reviewing the major questions posed by the study. This part of the internal rater training lasted forty-five minutes.

After a short break, the researcher followed a multi-step process to familiarize the raters with the task of scoring student responses. This "norming"



process was modeled on recommendations made by Maki.<sup>1261</sup> Maki refers to this process as “calibration” and describes calibration as the process of “establishing interrater reliability in scoring students texts”.<sup>1262</sup> She notes that calibration is “developed over successive applications of a scoring rubric to student work...to sure that rater responses are consistent.”<sup>1263</sup> Maki outlines the six steps in this process:

(1) Ask raters to independently score a set of student samples that reflects the range of texts students produce in response to a direct method. (2) Bring raters together to review their responses to identify patterns of consistent and inconsistent responses. (3) Discuss and then reconcile inconsistent responses. (4) Repeat the process of independent scoring on a new set of student samples. (5) Again, bring all scorers together to review their responses to identify patterns of consistent and inconsistent responses. (6) Discuss and then reconcile inconsistent responses. This process is repeated until raters reach consensus about applying the scoring rubric. Ordinarily, two to three of these sessions calibrate raters’ responses.<sup>1264</sup>

In this study, the researcher began by sharing 5 “anchor” responses to demonstrate the range of student responses with the raters and model the scoring process by “thinking aloud”. Then, the raters scored 5 more anchor responses independently, and then discussed the scores they assigned in their small groups. In discussions, raters were asked to focus on inconsistent scores and attempt to reconcile them. Next, the small groups reported their scores to the full group and the full group discussed the remaining inconsistencies and

---

<sup>1261</sup> Maki, Assessing for Learning. 127.

<sup>1262</sup> Maki, Assessing for Learning. 126.

<sup>1263</sup> Maki, Assessing for Learning. 126.

<sup>1264</sup> Maki, Assessing for Learning. 127.

attempted to reconcile them. This part of the training lasted seventy-five minutes. After a break for lunch, raters again scored five anchor responses independently, then discussed them in small groups and finally the full group attempting to reconcile inconsistencies in scoring. This time, the process took about sixty minutes.

Raters felt at this point that they were ready to score student responses on their own. As practice, each rater scored 10 student responses. These practice scores were not discussed with other raters, but were used instead by the researcher to explore the level of proficiency of each rater at the start of the study. After turning in their “practice” responses, raters began to score the 75 study responses. They received three packets of 25 student responses; as they turned in each packet, they received a new one for scoring. After raters scored all study responses, they completed the open-ended comment sheet and left the scoring session.

### ***The External Rater Experience***

The ten external (non-NCSU) raters did not participate in a training session. Instead they were provided with study materials, background information, and directions via the mail. When raters opened their study packets, they encountered several documents. The first document inventoried the contents of the study packets. The second document included the study title, the purpose of the study, the major research questions, information explaining their role in the study, and the directions for participating in the study. Raters were

directed to read and return their consent forms, rater data forms, and the open-ended comment sheet. They were made aware of the supporting materials offered in the packet including the URL for LOBO and directions to login as a guest, screenshots from LOBO, the full and student versions of the rubric, LOBO Information Literacy Skills Objectives and Outcomes, and the ACRL Information Literacy Competency Standards for Higher Education. The packet also included the 10 practice responses and the 75 study responses. Both were labeled with directions. Finally, raters were directed to place completed study materials in the postage-paid envelope and return them to the researcher.

### ***Preparation for Statistical Analysis***

At the close of the study, all raters returned their consent forms, data sheets, open-ended comment sheets, and rubric score sheets. The consent forms were filed for future reference, and the data sheets were used to summarize the rater populations. The open-ended comment sheets were transcribed for later use as a source for raters' perceptions and anecdotal comments. The rubric score sheets were organized for data entry and analysis. The data from the rubric score sheets was entered into an Excel spreadsheet. The number of the response, the position of the response among the 75 study responses, and the rater's number were included in the spreadsheet. For each response, each rater's score for the four criteria were recorded, along with the total score (0-8), and a grade equivalent (A, B, C, U) suggested by Mertler.<sup>1265</sup>

---

<sup>1265</sup> Mertler, "Designing Scoring Rubrics."

## Statistical Analysis of Reliability

Indeed, there is “nearly universal agreement” that reliability is an important property in educational measurement.<sup>1266</sup> Moskalko and other researchers describe reliability as a measure of consistency.<sup>1267</sup> Williams defines reliability as the concept that “the measurement can be repeated.”<sup>1268</sup> Other researchers suggest that reliability is more than getting the same score twice; they note that reliability in performance assessment measures adds the issue of rater/scorer consistency.<sup>1269</sup> Educational measurement experts describe multiple types of reliability, such as stability reliability, internal consistency reliability, alternate-form reliability,<sup>1270</sup> split half reliability, and rational equivalence reliability.<sup>1271</sup> However, Moskalko and Leydens suggest that these are “statistical methods that are used to establish consistency of student performances within a given test or across more than one test. These types of reliability are of more concern on standardized or high stakes testing than they are in classroom assessment.”<sup>1272</sup> Moskalko and Leydens state, “The two forms of reliability that typically are considered in

---

<sup>1266</sup> Colton, Gao, Harris, Kolen, Martinovich-Barhite, Wang and Welch, Reliability Issues with Performance Assessments. 3.

<sup>1267</sup> Moskalko, "Scoring Rubrics: What, When, and How?."

<sup>1268</sup> Williams, "Creativity in Assessment." 323.

<sup>1269</sup> Colton, Gao, Harris, Kolen, Martinovich-Barhite, Wang and Welch, Reliability Issues with Performance Assessments. 4.

<sup>1270</sup> Popham, Test Better, Teach Better. 52-53.

<sup>1271</sup> Moskalko, "Scoring Rubrics: What, When, and How?."

<sup>1272</sup> Moskalko, "Scoring Rubrics: What, When, and How?."

classroom assessment and in rubric development involve rater (or scorer) reliability.”<sup>1273</sup> The first form is interrater reliability, which refers to the consistency of scores assigned by multiple raters.<sup>1274</sup> The second is intrarater reliability, which refers to the consistency of scores assigned by one rater at different points of time.<sup>1275</sup> These reliability scores can be influenced by several factors, including “the objectivity of the task/item/scoring, the difficulty of the task/item, the group homogeneity of the examinees/raters, speededness, number of tasks/items/raters, and the domain coverage.”<sup>1276</sup> While these influencing factors must be considered, researchers insist, “how one intends to use an assessment should determine which type of reliability estimate is of most interest.”<sup>1277</sup> Because the rubric in this study could be used by both librarians and English 101 instructors to grade student responses, interrater reliability measures are of more concern than intrarater reliability.

### ***Interrater Reliability***

Many assessment methods require raters to judge or quantify some

---

<sup>1273</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1274</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1275</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1276</sup> Colton, Gao, Harris, Kolen, Martinovich-Barhite, Wang and Welch, Reliability Issues with Performance Assessments. 4-5.

<sup>1277</sup> Colton, Gao, Harris, Kolen, Martinovich-Barhite, Wang and Welch, Reliability Issues with Performance Assessments. 5.

aspect of student behavior.<sup>1278</sup> For example, raters are often used to “empirically test the viability of a new scoring rubric.”<sup>1279</sup> In such cases, interrater reliability is a very useful measure. Stemler states, “Interrater reliability is one of the most important concepts in educational and psychological measurement. Without demonstrating that two independent judges can be reliably trained to rate a particular behavior, our hope for achieving objective measurement of behavioral phenomena is diminished.”<sup>1280</sup> Johnson, Penny and Gordon challenge those who design and implement assessments, especially assessments using constructed-response items, to strive to achieve high levels of interrater reliability,<sup>1281</sup> and Stemler notes that achieving an acceptable level of interrater reliability impacts the validity of assessment results.<sup>1282</sup>

Interrater reliability refers to “the level of agreement between a particular set of judges on a particular instrument at a particular time” and “provide[s] a statistical estimate of the extent to which two or more judges are applying their ratings in a manner that is predictable and reliable.”<sup>1283</sup> According to Stemler, raters, or judges, are used when student products or performances cannot be

---

<sup>1278</sup> Steven E. Stemler, "A Comparison of Consensus, Consistency, and Measurement Approaches to Estimating Interrater Reliability," Practical Assessment, Research, and Evaluation 9.4 (2004).

<sup>1279</sup> Stemler, "A Comparison of Consensus."

<sup>1280</sup> Stemler, "A Comparison of Consensus."

<sup>1281</sup> Johnson, Penny and Gordon, "The Relation between Score Resolution Methods and Interrater Reliability." 123.

<sup>1282</sup> Stemler, "A Comparison of Consensus."

<sup>1283</sup> Stemler, "A Comparison of Consensus."

scored objectively as right or wrong, but require a rating of degree.<sup>1284</sup> This use of raters results in the subjectivity that comes hand in hand with a rater's interpretation of the product or performance.<sup>1285</sup> Johnson, Penny, and Gordon note that raters' subjective decisions "contribute a unique source of measurement error."<sup>1286</sup> They note that "if the reliability of rater decisions is low, then there is the likelihood that if a decision were to be rendered by another judge, the decision would differ. Such instability in a decision making process is often seen as a liability...[and] such a process is simply unfair" to those who are assessed.<sup>1287</sup>

In order to combat potential subjectivity and unfairness, many assessors develop rubrics to improve the interrater reliability of constructed-response and performance assessments. Moskal and Leydens state that rubrics respond to concerns of subjectivity and unfairness by formalizing the criteria for scoring a student product or performance.<sup>1288</sup> They write, "The descriptions of the score levels are used to guide the evaluation process. Although scoring rubrics do not completely eliminate variations between raters, a well-designed scoring rubric can reduce the occurrence of these discrepancies."<sup>1289</sup> In fact, Colton and his

---

<sup>1284</sup> Stemler, "A Comparison of Consensus."

<sup>1285</sup> Stemler, "A Comparison of Consensus."

<sup>1286</sup> Johnson, Penny and Gordon, "The Relation between Score Resolution Methods and Interrater Reliability."

<sup>1287</sup> Johnson, Penny and Gordon, "The Relation between Score Resolution Methods and Interrater Reliability."

<sup>1288</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1289</sup> Moskal, "Scoring Rubrics: What, When, and How?."

colleagues state, "Generally it has been found that it is possible to define rubrics so well that raters can be trained to score reliably."<sup>1290</sup> Wolfe, Koa and Ranney explain that raters can be trained to "abandon previously held values and to adopt those espoused by the rubric writers."<sup>1291</sup> They suggest that "initial small differences are likely to diminish even further because scorers undergo extensive training...to indoctrinate scorers into a system in which they become like-minded."<sup>1292</sup> Wolfe, Koa and Ranney also note that training raters with anchor papers creates an environment in which "it seems unlikely that even minimally proficient scorers would dramatically diverge from the scoring rubric in the weighting of scoring-focus categories when making scoring decisions."<sup>1293</sup>

In cases where raters using rubrics produce inconsistent scores, several problems may exist. Rater inconsistency can be due to inadequate training of raters, inadequate detail of rubrics, or "the inability of raters to internalize the rubrics."<sup>1294</sup> Wolfe, Koa and Ranney also suggest that "scorers with different levels of scoring ability do not focus on different [product or performance] features, but probably have different levels of understanding about the scoring criteria."<sup>1295</sup> Moskal and Leydens note that discussing rater scoring differences

---

<sup>1290</sup> Colton, Gao, Harris, Kolen, Martinovich-Barhite, Wang and Welch, Reliability Issues with Performance Assessments.

<sup>1291</sup> Edward W. Wolfe, Chi-Wen Kao and Michael Ranney, "Cognitive Differences in Proficient and Nonproficient Essay Scorers," Written Communication 15.4 (1998).

<sup>1292</sup> Wolfe, Kao and Ranney, "Cognitive Differences."

<sup>1293</sup> Wolfe, Kao and Ranney, "Cognitive Differences."

<sup>1294</sup> Colton, Gao, Harris, Kolen, Martinovich-Barhite, Wang and Welch, Reliability Issues with Performance Assessments. 9.

<sup>1295</sup> Wolfe, Kao and Ranney, "Cognitive Differences."



and making appropriate changes to rubrics is worthwhile and ultimately helps improve assessment reliability.<sup>1296</sup> Maki concurs and provides a six-step process to train raters to provide consistent and reliable scores.<sup>1297</sup> It is Maki's six-step process that forms the basis for the procedures followed in this study.

A second way to improve rater consistency is to make rubric criteria and performance descriptions more specific. However, some researchers warn that including rigid definitions in rubrics might limit their generalizability.<sup>1298</sup> Rubric generalizability is an important issue that impacts how results from one assessment of student learning might generalize to similar student learning scenarios.<sup>1299</sup> Moskal and Leydens describe the reliability concerns related to matching rubrics to student populations:

A scoring rubric that consistently measures the performances of one set of students may not consistently measure the performances of a different set of students. For example, if a task is embedded within a context, one population of students may be familiar with that context and the other population may be unfamiliar with that context. The students who are unfamiliar with the given context may achieve a lower score based on their lack of knowledge of the context. If these same students had completed a different task that covered the same material that was embedded in a familiar context, their scores may have been higher. When the cause to variation in performance and the resulting scores is unrelated to the purpose of the assessment, the scores are unreliable.<sup>1300</sup>

---

<sup>1296</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1297</sup> Bernier, "Making Yourself Indispensable." 126-127.

<sup>1298</sup> Colton, Gao, Harris, Kolen, Martinovich-Barhite, Wang and Welch, Reliability Issues with Performance Assessments. 5.

<sup>1299</sup> Colton, Gao, Harris, Kolen, Martinovich-Barhite, Wang and Welch, Reliability Issues with Performance Assessments. 9.

<sup>1300</sup> Moskal, "Scoring Rubrics: What, When, and How?."

Stemler adds, "Interrater reliability must be demonstrated anew for each study, even if the study is using a scoring rubric or instrument that has been shown to have high interrater reliability in the past."<sup>1301</sup> He reminds assessors that interrater reliability is a function of an assessment situation, not of an assessment tool itself.<sup>1302</sup>

Stemler also points out that there are three general categories of interrater reliability and that, in any study, a researcher should decide which type of interrater reliability best suits the purpose of the assessment and then be sure to treat and summarize the assessment data in ways that are consistent with that type of interrater reliability.<sup>1303</sup> The three types of interrater reliability are consensus estimates, consistency estimates, and measurement estimates.<sup>1304</sup> Consensus estimates are based on the belief that "reasonable observers should be able to come to exact agreement about how to apply the various levels of a scoring rubric to the observed behaviors."<sup>1305</sup> In contrast, consistency estimates are based on the assumption that "it is not really necessary for two judges to share a common meaning of the rating scale, so long as each judge is consistent in classifying the phenomenon according to his or her own definition of the scale."<sup>1306</sup> Finally, measurement estimates are based on the belief that "one

---

<sup>1301</sup> Stemler, "A Comparison of Consensus."

<sup>1302</sup> Stemler, "A Comparison of Consensus."

<sup>1303</sup> Stemler, "A Comparison of Consensus."

<sup>1304</sup> Stemler, "A Comparison of Consensus."

<sup>1305</sup> Stemler, "A Comparison of Consensus."

<sup>1306</sup> Stemler, "A Comparison of Consensus."

should use all of the information available from all judges (including discrepant ratings) when attempting to create a summary score for each respondent.”<sup>1307</sup>

Each of these types of interrater reliability carries with it “different implications for how data from multiple judges should be summarized most appropriately.”<sup>1308</sup>

For the focus of this study, consensus estimates are the most relevant form of interrater reliability.

### ***Consensus Estimates***

Consensus estimates of interrater reliability assume that independent raters should be able to agree on how to use a rubric to score student products or performances.<sup>1309</sup> If two raters can agree exactly on a rubric score to assign to a student’s work, then the two raters “may be said to share a common interpretation of the construct.”<sup>1310</sup> This type of estimate is most useful, according to Stemler, when data is “nominal in nature and different levels of the rating scale represent qualitatively different ideas.”<sup>1311</sup> He also says that consensus estimates are useful when “different levels of the rating scale are assumed to represent a linear continuum of the construct, but are ordinal in nature (e.g., a Likert scale). In that case, the judges must come to exact agreement about the quantitative levels of the construct under investigation,

---

<sup>1307</sup> Stemler, "A Comparison of Consensus."

<sup>1308</sup> Stemler, "A Comparison of Consensus."

<sup>1309</sup> Stemler, "A Comparison of Consensus."

<sup>1310</sup> Stemler, "A Comparison of Consensus."

<sup>1311</sup> Stemler, "A Comparison of Consensus."

rather than attempting to evaluate qualitative differences in scoring categories.”<sup>1312</sup>

There are three main ways of calculating consensus estimates of interrater reliability. The most popular method is the simple percent-agreement figure.<sup>1313</sup> This figure is calculated by “adding up the number of cases that received the same rating by both judges and dividing that number by the total number of cases rated by the two judges.”<sup>1314</sup> Three advantages of the simple percent-agreement statistic are that it has “strong intuitive appeal,” a simple calculation process, and it is easy to explain.<sup>1315</sup> There are also two disadvantages to this statistic. First, this calculation is used to compare two raters, and the present study includes twenty-five raters plus the researcher. Second, the percent-agreement statistic does not correct for chance. In other words, the statistic does not consider the random probability of a rater assigning a particular score. In rubric assessment, the limited amount of criteria and levels of performance description increase the probability of a rater assigning a particular score by chance, rather than intention. As a result, the percent-agreement statistic is likely to be artificially inflated. To correct for chance, there is a procedure to modify the percent-agreement statistic.<sup>1316</sup> The modification involves requiring not only exact agreement, but also adjacent scoring categories

---

<sup>1312</sup> Stemler, "A Comparison of Consensus."

<sup>1313</sup> Stemler, "A Comparison of Consensus."

<sup>1314</sup> Stemler, "A Comparison of Consensus."

<sup>1315</sup> Stemler, "A Comparison of Consensus."

<sup>1316</sup> Stemler, "A Comparison of Consensus."

on the rating scale.<sup>1317</sup> This relaxes the need for exact agreement among raters, but it has one disadvantage. If the rating scale has only a limited number of categories (such as a 1-4 scale), the estimate may be inflated.<sup>1318</sup> Stemler writes, "If the rating scale has a limited number of points, then nearly all points will be adjacent, and it would be surprising to find agreement lower than 90%. The technique of using adjacent categories results in a situation where the percent agreement at the extreme ends of the rating scale is almost always lower than the middle."<sup>1319</sup> Because the rubric used in this study had only three levels of performance description, this method was not used to analyze interrater reliability.

A second method of calculating a consensus estimate of interrater reliability is Kendall's coefficient of concordance. Kendall's coefficient is used to estimate agreement among multiple raters, corrects for chance, and is appropriate for ordinal responses that are numerically coded.<sup>1320</sup> Because the rubric used in this study yields responses that are both ordinal and numerically coded, Kendall's coefficient seems a good match for this study. However, one major disadvantage of this statistic is that it offers no agreed-upon index for interpreting results. That is, there are no cut-offs for levels of acceptable or

---

<sup>1317</sup> Stemler, "A Comparison of Consensus."

<sup>1318</sup> Stemler, "A Comparison of Consensus."

<sup>1319</sup> Stemler, "A Comparison of Consensus."

<sup>1320</sup> SAS, Compute Estimates and Tests of Agreement among Multiple Raters, 2006, 5 Jan 2006 <<http://support.sas.com/ctx/samples/index.jsp?sid=507&tab=details>>.

unacceptable reliability estimates. As a result, this statistic was not used to estimate interrater reliability.

The third method of calculating a consensus estimate of interrater reliability, and the method that used throughout in this study, is Cohen's kappa statistic.<sup>1321</sup> This statistic estimates the degree of consensus among multiple raters on nominal data after correcting for the "amount of agreement that could be expected by chance alone based on the values of the marginal distributions."<sup>1322</sup> Therefore, Cohen's kappa indicates whether the agreement among raters is better than chance would predict. Stemler explains:

The interpretation of the kappa statistic is slightly different than the interpretation of the percent-agreement figure. A value of zero on kappa does not indicate that the two judges did not agree at all; rather, it indicates that the two judges did not agree with each other any more than would be predicted by chance alone. Consequently, it is possible to have negative values of kappa if judges agree less often than chance would predict.<sup>1323</sup>

Furthermore, this statistic offers the advantage of an index that allows researchers to easily interpret results. Landis and Koch assign the labels found in Figure 5.3 to corresponding ranges of Cohen's kappa.<sup>1324</sup> Statistical support documentation points to this as the definitive index for kappa.<sup>1325</sup> As a final advantage, Stemler notes that "kappa is a highly useful statistic when one is

---

<sup>1321</sup> Stemler, "A Comparison of Consensus."

<sup>1322</sup> Stemler, "A Comparison of Consensus."

<sup>1323</sup> Stemler, "A Comparison of Consensus."

<sup>1324</sup> J. Richard Landis and Gary G. Koch, "The Measure of Observer Agreement for Categorical Data," *Biometrics* 33 (1977). 165.

<sup>1325</sup> SAS, Compute Estimates and Tests of Agreement.

concerned that the percent-agreement statistic may be artificially inflated due to the fact that most observations fall into a single category.<sup>1326</sup>

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 5.3. Kappa Statistics and Strength of Agreement*

There are two limitations of Cohen's kappa statistic. First, according to Stemler, "kappa values for different items or from different studies cannot be meaningfully compared unless the base rates are identical."<sup>1327</sup> Therefore, it is difficult to compare kappa statistics over different assessment situations."<sup>1328</sup> However, this is not a disadvantage that is significant in this study. Second, kappa requires a greater number of observances to achieve an acceptable standard error. This requirement is not significant for the major portion of this study involving seventy-five student responses. However, the practice test included in this study, which analyzes only ten student responses, is not well suited for analysis using the kappa statistic. Instead, the practice test was analyzed using the two-way cross-classification tables generated in SAS when the kappa statistic is run. These cross-classification tables provide anecdotal information, but cannot yield data of statistical significance.

<sup>1326</sup> Stemler, "A Comparison of Consensus."

<sup>1327</sup> Stemler, "A Comparison of Consensus."

<sup>1328</sup> Stemler, "A Comparison of Consensus."

There are several advantages to using consensus estimates. For instance, consensus estimates are well suited to working with “nominal variables whose levels on the rating scale represent qualitatively different categories.”<sup>1329</sup> Consensus estimates can also help determine how judges might misinterpret how to apply a rubric. Stemler states, “A visual analysis of the output allows the researcher to go back to the data and clarify the discrepancy or retain the judges.”<sup>1330</sup> Another exciting advantage of consensus estimates is that they identify raters who have been trained enough to agree on how to interpret a rating scale. When that occurs, two raters may be treated as equivalent, and both raters need not score all student products or performances. Stemler confirms:

When judges exhibit a high level of consensus, it implies that both judges are essentially providing the same information. One implication of a high consensus estimate of interrater reliability is that both judges need not score all remaining items...because the two judges have empirically demonstrated that they share a similar meaning for the scoring rubric. In practice, however, it is usually a good idea to build in a 30% overlap between judges even after they have been trained in order to provide evidence that the judges are not drifting from their consensus as they read more items.<sup>1331</sup>

When raters are trained to a level of agreement, summary scores can be figured by taking the score of one rater or averaging the scores given by all raters.<sup>1332</sup> Although this advantage is not explored in this study, it has practical implications for the application of rubrics in the future.

---

<sup>1329</sup> Stemler, "A Comparison of Consensus."

<sup>1330</sup> Stemler, "A Comparison of Consensus."

<sup>1331</sup> Stemler, "A Comparison of Consensus."

<sup>1332</sup> Stemler, "A Comparison of Consensus."



## Analysis of Validity

Before examining the validity of an assessment tool, its reliability must be confirmed.<sup>1333</sup> If an assessment is unreliable, then it will not provide scores that can be used to draw valid inferences.<sup>1334</sup> Once score consistency is determined, then assessors should turn to the content of a test and ensure that it fits the intended purposes of assessment.<sup>1335</sup>

The term “validity” is used to refer to the meaningfulness of an assessment measure.<sup>1336</sup> Assessors must determine whether the tool measures “what it is intended to measure.”<sup>1337</sup> Prus and Johnson contend that validity is the “key selection criterion” for any higher education assessment.<sup>1338</sup> Validity is made up of three attributes including relevance, accuracy, and utility.<sup>1339</sup> A relevant assessment measures outcomes as directly as possible, an accurate assessment measures outcomes as precisely as possible, and a utilitarian assessment measures outcomes both formatively and summatively and provides

---

<sup>1333</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment. 10.

<sup>1334</sup> Popham, Test Better, Teach Better. 54.

<sup>1335</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment. 10.

<sup>1336</sup> Grassian and Kaplowitz, Information Literacy Instruction. 275.

<sup>1337</sup> Grassian and Kaplowitz, Information Literacy Instruction. 275.

<sup>1338</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 70.

<sup>1339</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 70.

results that can be applied to program evaluation and improvement.<sup>1340</sup> Moskal and Leydens suggest that more than one form of validity evidence might need to be considered if an assessment measure is intended to serve multiple purposes.<sup>1341</sup> Two types of validity evidence that are relevant to this study are content-related evidence and construct-related evidence.<sup>1342</sup>

### ***Content-Related Validity***

Assessors who design a measure to “elicit evidence of an individual's knowledge within a given content area” should consider content-related evidence.<sup>1343</sup> Content-related evidence is used to establish whether or not a “test's items satisfactorily reflect the content the test is supposed to represent.”<sup>1344</sup> The main method of measuring content-related validity is to use human judgment.<sup>1345</sup> Experts should be able to examine a test and deem the questions “reasonable”.<sup>1346</sup> Moskal and Leydens suggest that content-related validity is especially important to consider when developing scoring rubrics.<sup>1347</sup> They suggest that content-related validity refers to both “the extent to which a

---

<sup>1340</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 70.

<sup>1341</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1342</sup> Popham, Test Better, Teach Better. 45.

<sup>1343</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1344</sup> Popham, Test Better, Teach Better. 47.

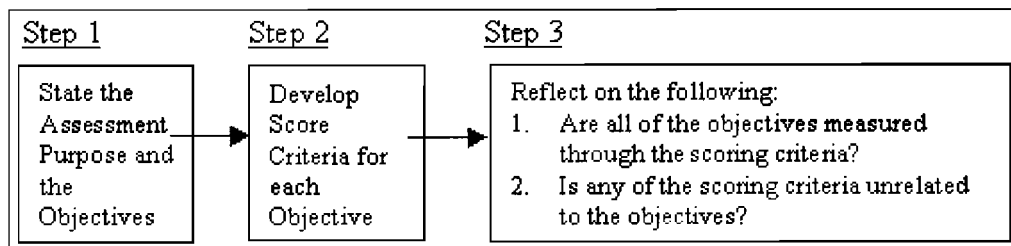
<sup>1345</sup> Popham, Test Better, Teach Better. 47.

<sup>1346</sup> Williams, "Creativity in Assessment." 323.

<sup>1347</sup> Moskal, "Scoring Rubrics: What, When, and How?."

student's responses to a given assessment instrument reflects that student's knowledge of the content area...and the extent to which the assessment instrument adequately samples the content domain."<sup>1348</sup> In order to make clear the attributes of content-related valid test, Popham offers an example of an invalid test for comparison—a multiple-choice writing test. He writes, "The trouble was...all of my [writing skills] exams contained only multiple-choice items about the mechanics of writing. With suitable shame, I now confess that I never assessed my students' writing skills by asking them to write anything! What a cluck."<sup>1349</sup> Invalid assessments like the one Popham describes are dangerous in that they lead to incorrect conclusions that further lead to "unsound instructional decisions."<sup>1350</sup>

In order to ensure content-related validity of rubric assessments, Moskal and Leydens offer a three-step process (see Figure 5.4).<sup>1351</sup>



*Figure 5.4. Process to Ensure Content-Related Validity*

This three-step process was adhered to in this study. The purpose of the assessment was tied to the LOBO Information Literacy Skills Objectives and

<sup>1348</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1349</sup> Popham, *Test Better, Teach Better*. 51.

<sup>1350</sup> Popham, *Test Better, Teach Better*. 51.

<sup>1351</sup> Moskal, "Scoring Rubrics: What, When, and How?."

Outcomes, the ACRL Information Literacy Competency Standards for Higher Education, and the ACRL Objectives for Information Literacy Instruction: A Model Statement for Academic Librarians. The criteria that comprise the rubric were based on these same LOBO and ACRL learning outcomes.

Moskal and Leydens also suggest posing three questions to examine the content-related validity of a rubric: (1) Do the evaluation criteria address any extraneous content? (2) Do the evaluation criteria of the scoring rubric address all aspects of the intended content? (3) Is there any content addressed in the task that should be evaluated through the rubric, but is not?<sup>1352</sup> Librarians and NCSU assessment professionals examined the study rubric to check for content-related validity, and the rubric was revised based on their feedback prior to use in the study.

### ***Construct-Related Validity***

Assessors who design a measure to “capture reasoning, problem solving or other processes that are internal to the individual and, therefore, require more indirect examination” should consider construct-related validity.<sup>1353</sup> Construct-related validity “specifically addresses the questions of whether the test measures the trait, attribute, or mental process it is purported to measure.”<sup>1354</sup> Most researchers agree that construct-related validity is the most important and

---

<sup>1352</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1353</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1354</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment.

comprehensive type of validity because it encompasses content-related and criterion-related validity.<sup>1355</sup> Moskal and Leydens explain the importance of construct-related validity in rubric development:

Construct-related evidence is the evidence that supports that an assessment instrument is completely and only measuring the intended construct.... Regardless of the construct, an effort should be made to identify the facets of the construct that may be displayed and that would provide convincing evidence of the students' underlying processes. These facets should then be carefully considered in the development of the assessment instrument and in the establishment of scoring criteria.<sup>1356</sup>

Moskal and Leydens also identify two questions to examine the construct-validity of a rubric: (1) Are all of the important facets of the intended construct evaluated through the scoring criteria? (2) Are any of the evaluation criteria irrelevant to the construct of interest?<sup>1357</sup> By reflecting on the construct-related and content-related validity of a rubric during the drafting process, assessors can be more certain that they are measuring what they intend to measure. Such reflection also ensures that assessment results can be used to make decisions for improved learning experiences for students. Librarians and NCSU assessment professionals examined the study rubric to check for construct-related validity, and the rubric was revised based on their feedback prior to use in the study.

---

<sup>1355</sup> National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, The NPEC Sourcebook on Assessment.

<sup>1356</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1357</sup> Moskal, "Scoring Rubrics: What, When, and How?."

### ***A “Gold Standard” Test for Validity***

A statistical measure formed the primary method for investigating the validity of this study. For this study, Cohen’s kappa was used not only to explore interrater reliability, but also to explore the validity of this approach to information literacy assessment. While interrater reliability measures do not ensure the validity of an approach, it is an accepted practice to compare a group of raters to a “gold standard” to check for validity.<sup>1358</sup> Gwet explains that the gold standard is the “correct classification of subjects made by an experienced observer.”<sup>1359</sup> 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.”<sup>1360</sup> Gwet 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.”<sup>1361</sup> This approach is also known as a “rater-to-standard reliability” or “rater-to-expert reliability”.<sup>1362</sup> Using Cohen’s kappa, this study compared each rater’s scores to the gold standard set by the researcher. Then the raters were ranked according to their rater-to-standard reliability. In

---

<sup>1358</sup> Kilem Gwet, Handbook of Inter-Rater Reliability: How to Estimate the Level of Agreement between Two or Multiple Raters (Gaithersburg, Maryland: STATAXIS, 2001). 202.

<sup>1359</sup> Gwet, Handbook of Inter-Rater Reliability. 202.

<sup>1360</sup> Gwet, Handbook of Inter-Rater Reliability. 223.

<sup>1361</sup> Gwet, Handbook of Inter-Rater Reliability. 223.

<sup>1362</sup> Gwet, Handbook of Inter-Rater Reliability. 223.

estimating students' performance, ratings from the most "expert" or valid raters were used.

### **Funding**

The Committee on Undergraduate Program Review (CUPR) at NCSU provided funding for this study. Until 2006, CUPR awarded mini-grants to units on campus to encourage innovation in assessing undergraduate programs. Although the library instruction program at NCSU was not one of the programs that CUPR reviews, they acknowledged the importance of the partnership between the NCSU Libraries and the First-Year Writing Program by awarding this grant. This grant allowed the researcher to pay internal raters \$125 and external raters \$75 for participation in the study. Additional funding was secured from the School of Information and Library Science to pay external raters, mail study packets to external raters, and provide refreshments at the training session.

## **CHAPTER 6**

### **RESULTS**

By using a survey design methodology, this study analyzed the consistency with which rubric scores are assigned by raters in a variety of groups. It examined the validity of these scores by comparing them to the scores assigned by the researcher, and it used the scores to measure the achievement of information literacy learning outcomes. This chapter describes the interrater reliability of the rater groups, analyzes the validity of the scores assigned by raters, and summarizes the achievement of learning outcomes captured by the study. This chapter also provides data to answer the research questions that form the focus of the study: (1) To what degree can different groups of raters provide consistent scoring of artifacts of student learning using a rubric? (2) To what degree can raters provide scores that are consistent with scores assigned by the researcher? (3) To what degree are students able to use authority as a criterion to evaluate a website?

#### **Reliability of Rubric Assessment**

The major purpose of this study was to determine to what degree different groups of raters can provide consistent scoring of artifacts of student learning using a rubric. To this purpose, 25 raters scored 75 student responses from the LOBO tutorial as artifacts of student learning regarding the evaluation of website



authority. The reliability of scores was examined both within groups and across groups using Cohen's kappa statistics. This statistic was calculated for each of the four criteria included in the rubric. It was also calculated for the total score assigned to the student response. Because of limitations of the kappa statistic, total scores (0-8) were converted to letter grades (A, B, C, U) according to recommendations in the literature (see Figure 6.1).<sup>1363</sup>

Total Rubric Score	Letter Grade
7-8	A
5-6	B
3-4	C
0-2	U

*Figure 6.1. Total Score to Letter Grade Conversion*

After this conversion, Cohen's kappa was run on the "grade" assigned to the total student response as well.

To illustrate the reliability within each group of raters, charts were generated that show the kappa for each group of raters. To clarify the meaning of each kappa statistic, the level of agreement indicated by the kappa scores was color-coded using the index provided by Landis and Koch (see Figure 6.2).<sup>1364</sup> For example, rater groups that produce a kappa of .41-.60 are shaded blue according to the legend in Figure 6.2 to show that this kappa range corresponds to a moderate level of agreement among raters.

<sup>1363</sup> Mertler, "Designing Scoring Rubrics."

<sup>1364</sup> Landis and Koch, "The Measure of Observer Agreement." 165.

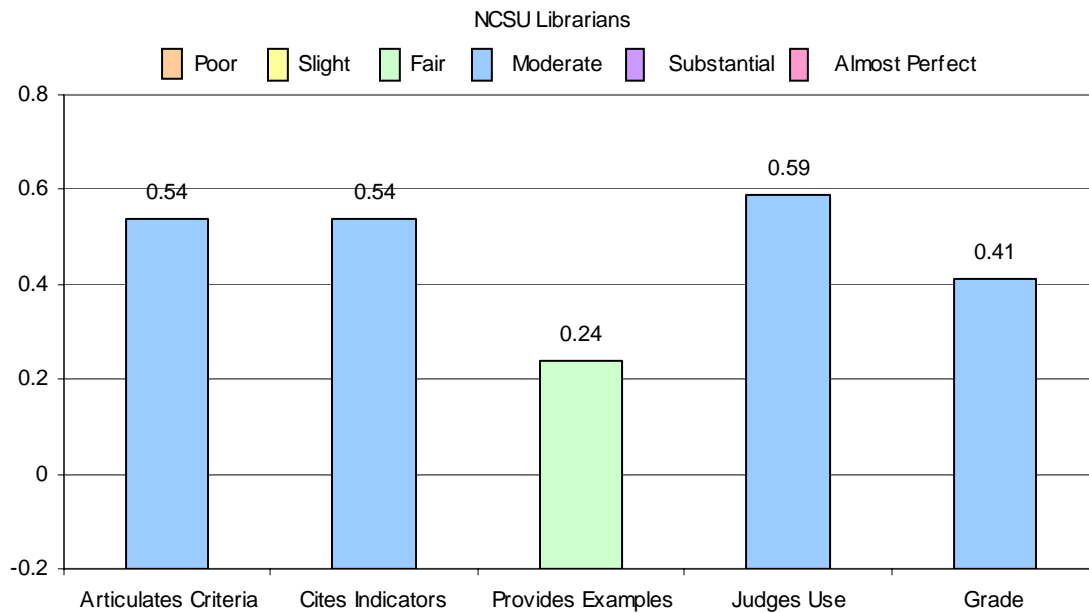
Kappa Statistic	Strength of Agreement
<0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

*Figure 6.2. Kappa Statistics and Strength of Agreement Color Codes*

### ***Reliability Within Rater Groups***

To determine the reliability of rubric scores provided by raters, each rater group was examined separately. The groups defined by the study are NCSU librarians, ENG 101 instructors, ENG 101 students, external (non-NCSU) instruction librarians, and external (non-NCSU) reference librarians. The raters' scores were also examined in two large groups: internal (NCSU) raters and external (non-NCSU) raters. Finally, the scores assigned by all raters were examined for consistency across all groups.

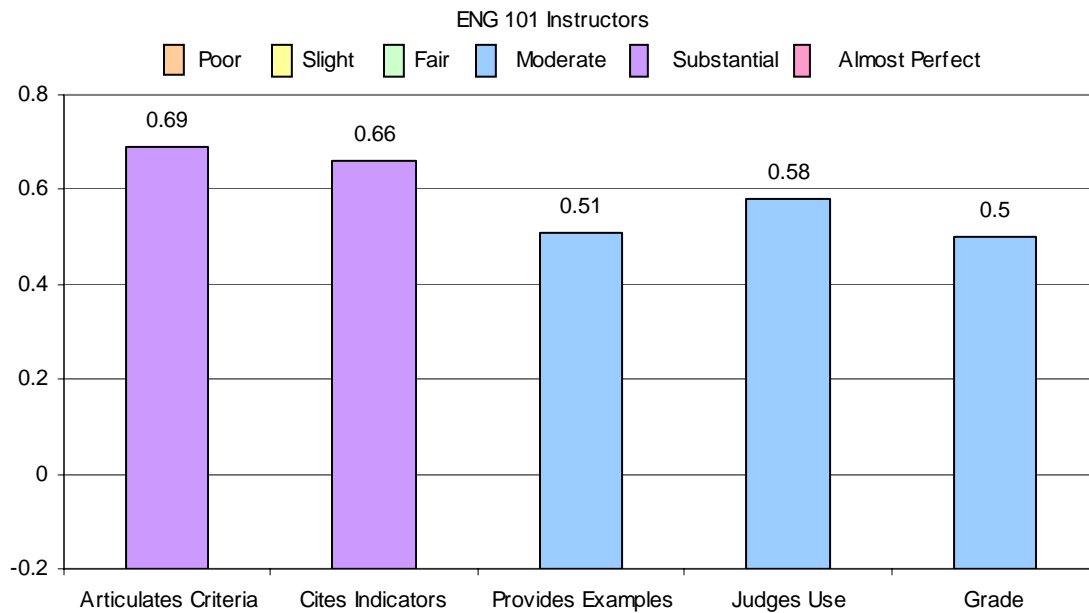
The first of five rater groups, NCSU librarians provided a moderate level of agreement when scoring student responses to the LOBO tutorial (see Figure 6.3). In three of the four criteria listed on the study rubric, NCSU librarians provided moderately reliable scores. NCSU librarians' ratings of the first rubric criterion, "Articulates Criteria", yielded a moderate kappa of .54 (standard error .03). On the second rubric criterion, "Cites Indicators of Criteria", their scores produced a moderate kappa of .54 as well (standard error .03). For the third rubric criterion, "Links Indicators to Example from Source", the ratings provided



*Figure 6.3. Kappa Statistics for NCSU Librarians*

by NCSU librarians showed a fair kappa of only .24 (standard error .03). Still, the librarians produced a moderate kappa of .59 (standard error .03) for the final criterion, “Judges Whether or Not to Use Source”. After the total rubric numerical scores were calculated and transformed into letter grades, a kappa statistic of .41 (standard error .02) shows that NCSU librarians produced moderate agreement on the grade they assigned to student responses. These kappa statistics indicate that, within their group, NCSU librarians provided moderately reliable scores, but had difficulty coming to consensus on the third criterion of the rubric, “Links Indicators to Examples from Source.”

As the second rater group, the ENG 101 instructors achieved moderately and substantially reliable results when scoring student responses to the LOBO tutorial (see Figure 6.4). For two of the criteria included in the study rubric,

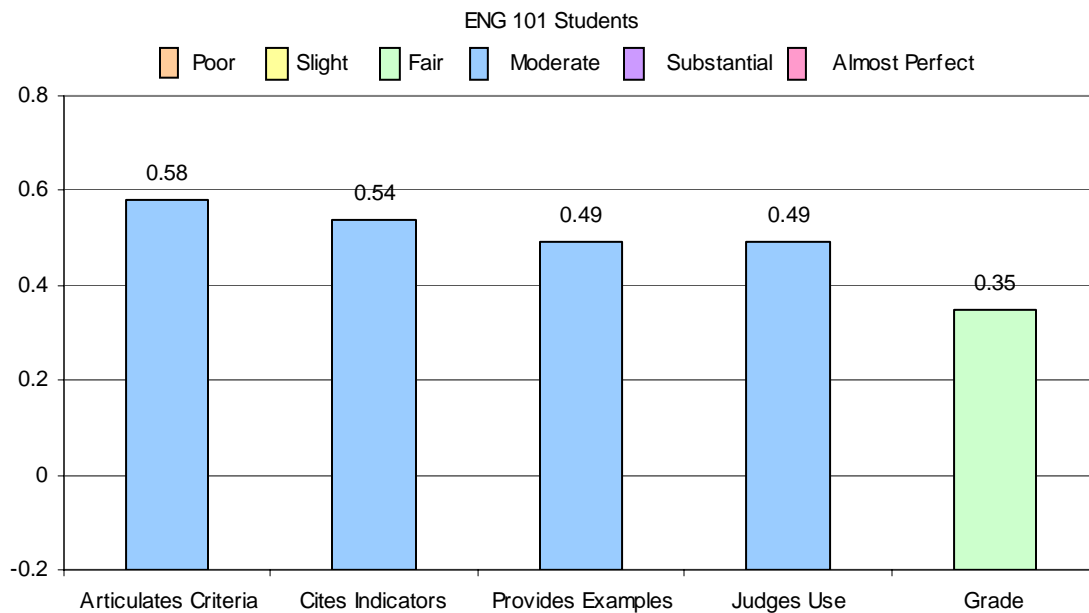


*Figure 6.4. Kappa Statistics for ENG 101 Instructors*

instructors produced substantially reliable scores. For the first criterion, “Articulates Criteria”, instructors’ ratings yielded a kappa of .69 (standard error .03), and for the second criterion, “Cites Indicators of Criteria”, they yielded a kappa of .66 (standard error .03). Both kappa scores correspond to a substantial level of agreement. For the third and fourth criteria on the study rubric, “Links Indicators to Examples from Source” and “Judges Whether or Not to Use Source”, the instructors provided rankings with kappas of .51 (standard error .03) and .58 (standard error .03) respectively, showing moderate levels of agreement. For the overall grade assigned to student responses, ENG 101 instructors’ ratings produced a kappa of .50 (standard error .03), indicating a moderate level of agreement. These kappas demonstrate that, within their rater group, ENG 101 instructors were able to provide moderately to substantially reliable scores in all areas of the rubric and in the total grade assigned to student responses. In fact,

ENG 101 instructors produced the greatest within group reliability of all rater groups studied.

The third rater group was comprised of ENG 101 students. This group produced a fair to moderate level of agreement with their scores of responses to the LOBO tutorial (see Figure 6.5). Interestingly, across the four criteria included



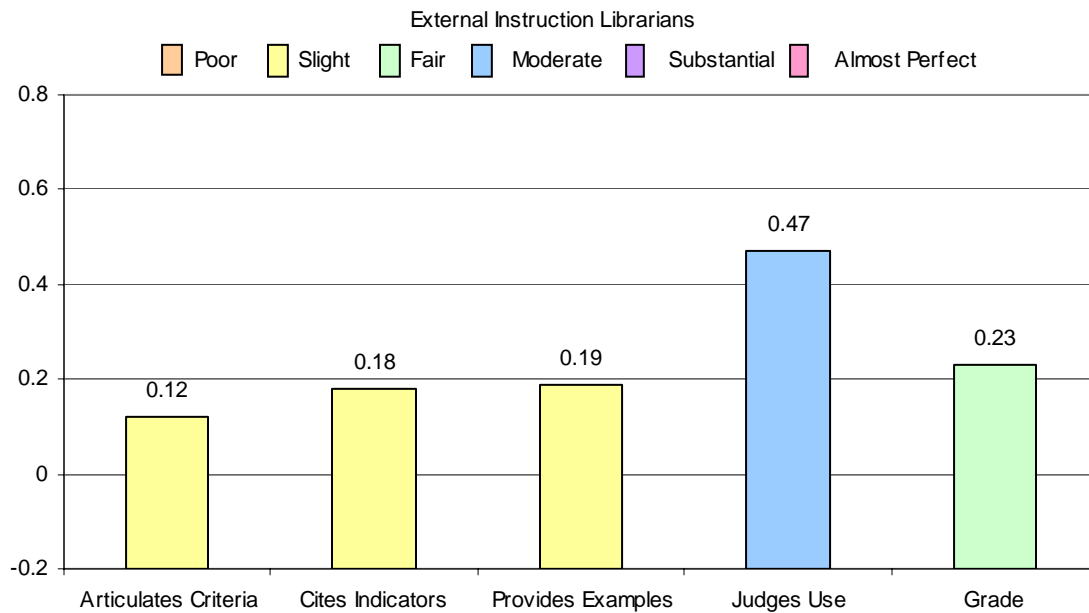
*Figure 6.5. Kappa Statistics for ENG 101 Students*

in the rubric, the students provided moderately reliable scores. For the first criterion, “Articulates Criteria”, the students’ scoring yields a kappa of .58 (standard error .03). For the second criterion, “Cites Indicators of Criteria”, the kappa for student scores is .54 (standard error .03). For both the third and fourth criteria, “Links Indicators to Examples from Source” and “Judges Whether or Not to Use Source”, the students’ ratings showed a kappa of .49 (standard error .03). Across all the criteria in the rubric, these kappa statistics indicate a moderate level of agreement. However, ENG 101 students’ ratings of the total response,

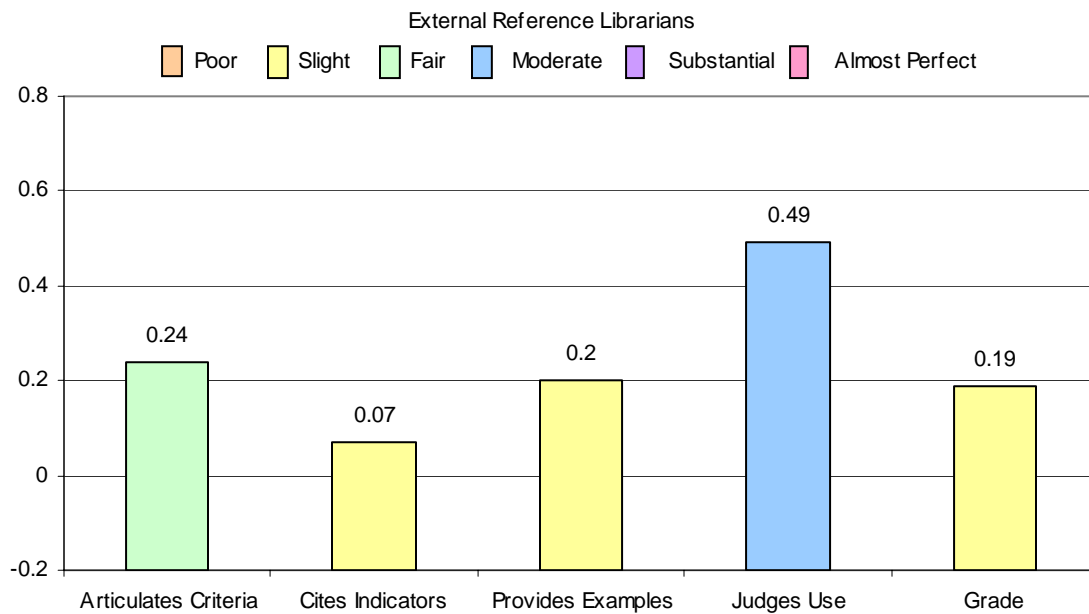
when converted to a letter score, indicate only a fair level of agreement with a kappa of .35 (standard error .03). These statistics indicate that, within their rater group, ENG 101 students were able to achieve moderate levels of reliability for each criterion included in the rubric, but only achieved a fair level of agreement on the grade assigned to student responses.

The last two rater groups included external instruction and reference librarians. These two groups provided scores that demonstrate slight and fair agreement (see Figures 6.6 and 6.7). While external instruction librarians achieved moderate agreement for the fourth criterion of the study rubric, “Judges Whether or Not to Use a Source”, with a kappa of .47 (standard error .03), their ratings for the other three criteria showed only slight agreement. Kappas for the first three criteria were .12, .18, and .19 (standard error .03). For the grade assigned to student responses, external instruction librarians demonstrated a fair level of agreement with a kappa of .23 (standard error .02). Overall, the levels of agreement produced by external instruction librarians were lower than would be acceptable.

External reference librarians also produced slight and fair levels of agreement. Like the external instruction librarians, external reference librarians came to moderate agreement on the fourth criteria of the rubric, “Judges Whether or Not to Use Source”, with a kappa of .49 (standard error .03). For the second and third rubric criteria, “Cites Indicators of Criteria” and “Links Indicators of Criteria”, external reference librarians produced slight levels of agreement with kappas of .07 and .20 (standard error .03) respectively. For the grade assigned



*Figure 6.6. Kappa Statistics for External Instruction Librarians*

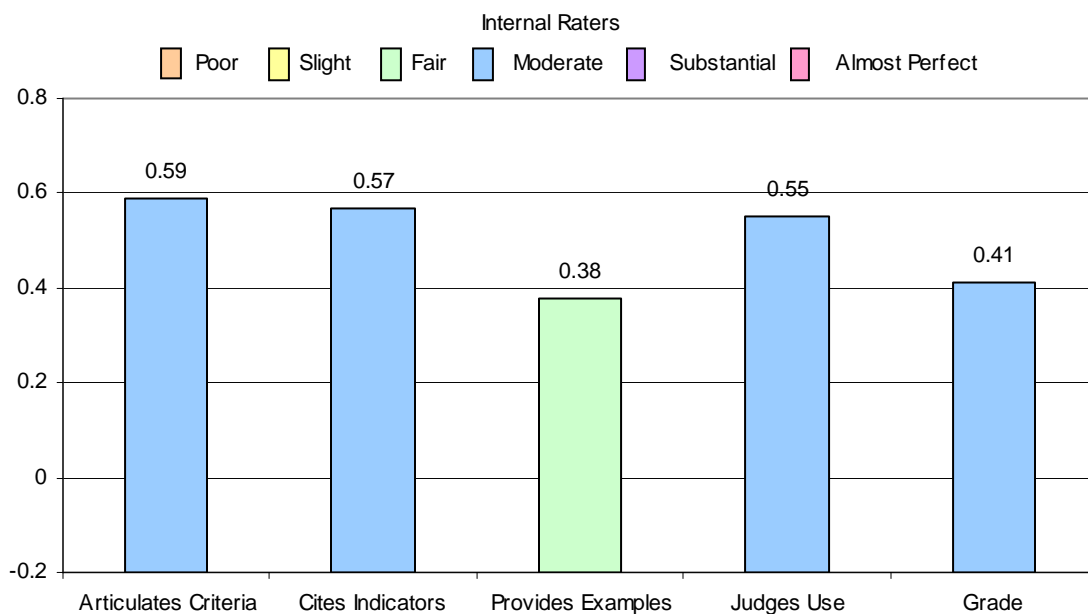


*Figure 6.7. Kappa Statistics for External Reference Librarians*

to the overall student responses, a kappa of .19 (standard error .02) shows that external reference librarians' ratings demonstrated only a slight level of

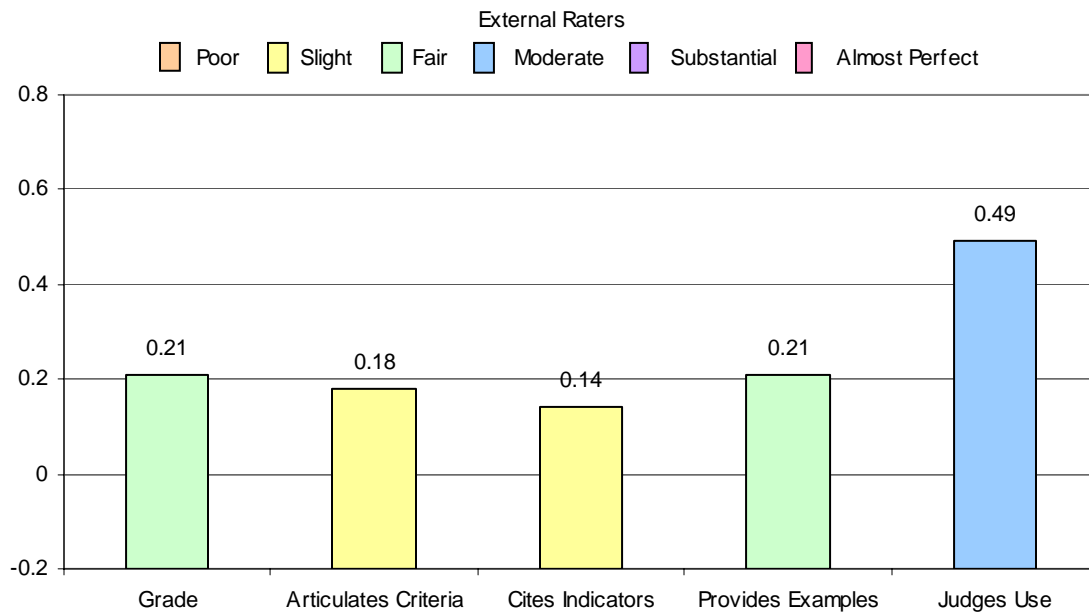
agreement. These kappa statistics show that, within their two groups, external reference and instruction librarians were unable to achieve greater than slight to fair levels of agreement in all areas except the fourth criterion of the rubric, “Judges Whether or Not to Use Source”. For this criterion alone, both groups of external librarians achieved a moderate level of agreement.

The scores assigned to student responses by the 25 raters in this study can be grouped differently into two larger categories: internal raters and external raters. Kappa statistics for these two larger categories can be seen in Figures 6.8 and 6.9. Overall, internal raters yielded moderate levels of agreement ranging from .55 to .59 (standard error .01) for the first, second, and fourth rubric criteria. For the third rubric criteria, the internal raters’ scores produced a kappa



*Figure 6.8. Kappa Statistics for Internal Raters*





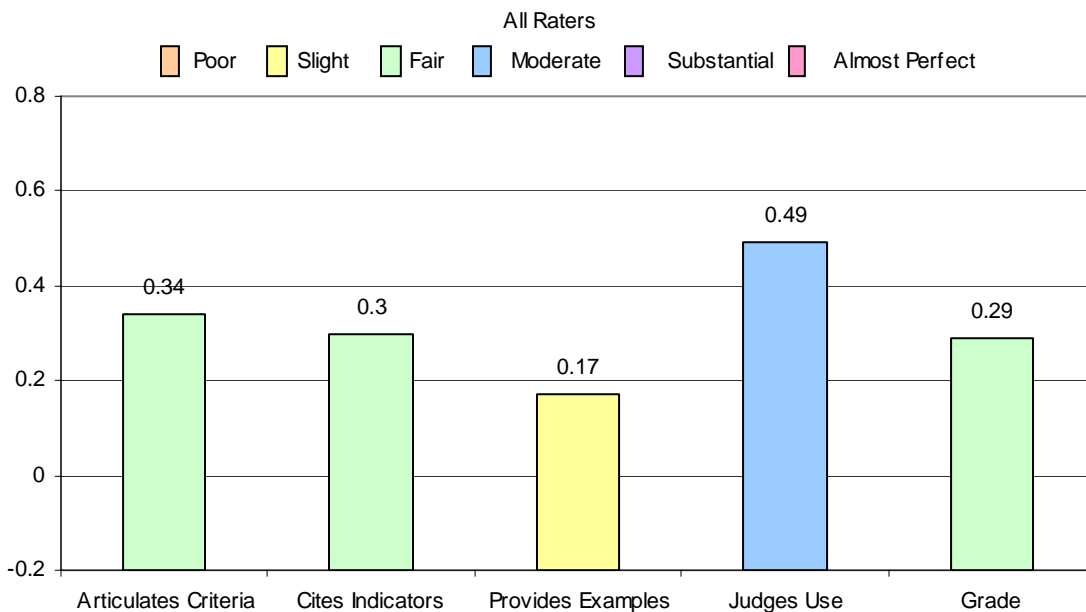
*Figure 6.9. Kappa Statistics for External Raters*

of .38, showing a fair level of agreement on this criterion. Internal raters produced a moderate level of agreement for the grades assigned to student responses. The kappa statistic for this measure was .41 (standard error .01). These levels demonstrate a generally moderate level of agreement within all internal raters. The third rubric criterion, “Links Indicators to Examples from Source”, produced only a fair kappa statistic.

In contrast, external raters produced only slight to fair levels of agreement. For the first and second criteria included in the rubric, external raters’ scores showed only slight agreement with kappa statistics of .18 and .14 (standard error .01) respectively. The third rubric criterion, “Links Indicators with Examples from Source”, shows a kappa of .21 (standard error .01), which indicates a fair level of agreement. The fourth criterion shows moderate agreement with a kappa of .49

(standard error .01), but the grade assigned by external raters indicates a fair level of agreement with a kappa of .21 (standard error .01).

As one large group, all 25 raters who participated in this study produced mostly fair levels of agreement (see Figure 6.10). For the first two criteria



*Figure 6.10. Kappa Statistics for All Raters*

included in the study rubric, “Articulates Criteria” and “Cites Indicators of Criteria”, raters agreed at a fair level, showing kappas of .34 and .30 (standard error .005) respectively. For the third rubric criterion, “Links Indicators to Examples from Source”, there was only slight agreement across all raters, demonstrated by a kappa of .17 (standard error .005). A moderate level of agreement was attained by all raters for the fourth criteria. However, only fair agreement was achieved for the grades assigned to student responses when all raters are grouped together. This is shown by a kappa of .29 (standard error .004). Overall, the large group including all raters did not achieve desirable levels of agreement.

### ***Significant Differences Among Rater Groups***

A number of statistically significant differences in reliability were revealed by analyzing the kappa statistics of rater groups using two-sided t-tests with an alpha level of .05. At the ninety-five percent confidence level, a t statistic over 1.96 is deemed significant. Statistically significant differences were found when comparing NCSU librarians with instructors, instructors with students, NCSU librarians with students, external reference librarians with external instruction librarians, NCSU librarians with external librarians, and all of the internal raters with all of the external raters.

Comparing the reliabilities of NCSU librarians and ENG 101 instructors revealed four significant differences between these two rater groups. The first statistically significant difference involved the first three criteria on the study rubric. For the first criterion, “Articulates Criteria”, the instructors’ substantial kappa level of .69 was significantly greater than the NCSU librarians’ moderate kappa of .54 ( $t = 3.5$ ), indicating that the ENG 101 instructors produced scores showing a greater degree of reliability for this rubric criterion. The second criterion of the rubric, “Cites Indicators of Criteria”, was scored with significantly greater reliability by instructors with a substantial kappa of .66 than the NCSU librarians with a moderate kappa of .54 ( $t = 2.8$ ), and the third criterion, “Links Indicators to Examples from Source”, was also scored with significantly greater reliability by instructors with a moderate kappa of .51 than the librarians with a fair kappa of .24 ( $t = 6.4$ ). Finally, the reliability of the grades assigned by

instructors with a moderate kappa of .50 was significantly greater than kappa produced by librarians, a .41 ( $t = 2.5$ ). There was no statistically significant difference between the reliability of the scores produced by NCSU librarians and ENG 101 instructors for the fourth criterion of the study rubric, “Judges Whether or Not to Use Source.” Taken as a whole, these significant differences indicate that the ENG 101 instructors produced more reliable scores of student responses than NCSU librarians.

There were four statistically significant differences in the reliabilities of the scores produced by ENG 101 instructors and those produced by ENG 101 students. In three areas of the study rubric, the scores supplied by instructors had significantly greater reliability than the students’. The first criterion of the rubric, “Articulates Criteria”, was more reliably scored by instructors with a substantial kappa of .69 than by students with a moderate kappa of .58 ( $t = 2.6$ ). The second criterion also revealed a statistically significant difference. For “Cites Indicators of Criteria”, instructors produced more reliable results, showing a substantial kappa of .66, than the students ( $t = 2.8$ ). For this criterion, the students’ ratings yielded only a moderate kappa of .54. Instructors were also shown to produce more reliable scores for the fourth criterion, “Judges Whether or Not to Use Source”, as well. This was indicated by a statistically significant difference ( $t = 2.1$ ) between the instructor’s kappa for this criterion at .58 and the students’ at .49. The fourth statistically significant difference ( $t = 3.5$ ) appeared when the moderate reliability of the grades assigned by the ENG 101 instructors ( $k = .50$ ) was compared to the fair reliability of the grades assigned by the

students ( $k = .35$ ). There was no statistically significant difference between the reliability of the scores produced by ENG 101 instructors and ENG 101 students for the third criterion of the rubric, "Links Indicators to Examples from Source." Overall, these significant differences signify that ENG 101 instructors scored student responses more reliably than ENG 101 students did.

The reliability of the scores assigned by NCSU librarians and ENG 101 students differed in only two statistically significant ways. For the third criterion of the study rubric, "Links Indicators to Examples from Source", the moderate kappa statistic of .49 provided by the students' ratings was significantly greater than the fair kappa for the librarians' ratings, which was only .24 ( $t = 5.9$ ). While both librarians and students showed moderately reliable scoring for the fourth rubric criterion, "Judges Whether or Not to Use Source", the librarians' scores ( $k = .59$ ) were significantly greater than the student's ( $k = .49$ ). The  $t$  statistic for this difference was 2.1. There was no significant difference between the reliability of librarians' scores and the reliability of the students' scores for the first rubric criterion, the second rubric criterion, and the grade assigned to student responses. The lack of significant differences between the NCSU librarians and ENG 101 students indicated that neither of the groups produced more reliable scores than the other. The fact that the two significant differences that appeared in the data were split (one shows the librarians with greater reliability, the other shows the students to produce more reliable results) underscores the lack of substantial differences between the scores of these two groups.

There were two statistically significant differences between the reliability of scores provided by external instruction librarians and external reference librarians. For the first criterion of the study rubric, external reference librarians showed significantly greater reliability ( $t = 2.8$ ), although the reliability of the external reference librarians on this criterion was still only fair. For the second criterion of the rubric, external instruction librarians demonstrated significantly greater reliability ( $t = 2.59$ ), but the reliability of both groups showed only a slight agreement. There were no significant differences between the reliability of the scores provided by external instruction librarians and external reference librarians for the third rubric criterion, the fourth rubric criterion, and the final grades assigned to the student responses. Because only two significant differences were identified between these two groups and these significant differences indicated greater reliability in opposing directions (one showing the greater reliability of external reference librarians and the other showing the greater reliability of external instruction librarians), these two groups did not appear to provide substantially different scores for student responses.

The reliability of the scores assigned by NCSU librarians differed significantly from the reliability of scores provided by external librarians in four ways. First, the kappa for the NCSU librarians' scores for the first criterion in the study rubric showed a moderate level of agreement ( $k = .54$ ). The kappa for external librarians was significantly lower at .18 showing only slight agreement ( $t = 11.4$ ). Similarly, the kappa for NCSU librarians' scores for the second rubric criterion was a moderate .54, while the external librarians showed only slight

agreement with a kappa of .14. These kappa statistics were significantly different ( $t = 12.6$ ). The t-test for the fourth rubric criteria showed a statistically significant difference ( $t = 3.2$ ) between the reliability of NCSU librarians' ratings ( $k = .59$ ) and those of the external librarians ( $k = .49$ ). The reliability of the grade assigned by the raters was also significantly different ( $t = 6.4$ ) with the NCSU librarians showing moderate reliability with a kappa of .41 and the external librarians demonstrating only fair reliability with a kappa of .21. Although there was no significant difference between these two groups for the third criterion of the study rubric, NCSU librarians clearly produced more reliable scores of student responses than their external counterparts.

In the same way, the reliability of all ratings for internal raters was significantly greater than all ratings for external raters. For the first criterion, "Articulates Criteria", the internal raters demonstrated a moderate kappa of .59 while the external raters provided only slight agreement ( $t = 29.0$ ). For the second rubric criterion, "Cites Indicators of Criteria", the kappa for internal raters was .57. This kappa is also significantly different ( $t = 30.4$ ) from the external raters ( $k = .14$ ). Likewise, internal raters produced greater reliability for the third ( $t = 12.0$ ) and fourth criteria ( $t = 4.2$ ). Finally, internal raters yielded moderate levels of agreement ( $k = .41$ ) when assigning grades to student response, a significant difference ( $t = 14.1$ ) from the fair level of agreement provided by external raters ( $k = .21$ ). The significantly greater level of agreement among internal raters over external raters was clearly demonstrated.

### ***Reliability Differences Throughout the Scoring Process***

One interesting finding was that the reliability of all raters increased as they scored the 75 student responses. When comparing the reliabilities of the first third, middle third, and last third of student responses using a Bonferroni adjustment, a t-test over 2.5 indicates a significant difference. In this study, the scores raters assigned to responses 26-50 were more reliable than the scores they assigned for responses 1-25 ( $t = 3.03$ ). Also, the scores assigned to responses 51-75 were more reliable than those assigned to 26-50 ( $t = 3.03$ ). It should be noted that, because the groups of responses compared in this test were scored by the same raters, these t-test scores are statistically conservative. As a result, a greater difference might actually exist than what these t-scores indicate. Because of standard error increases when only 25 responses are examined (rather than the full 75), analysis of the smaller rater groups was not statistically feasible.

### ***Expert Raters***

In addition to the rater groups described in the preceding section, analysis of a final rater group provided valuable insight into the reliability and validity of rubrics used to assess information literacy skills. While preparing the study data for statistical analysis, the researcher found that a subset of raters produced highly reliable and accurate scores. This subset did not align perfectly with any of the pre-established rater groups. Having anticipated the likelihood of an additional “expert” group of raters from the outset, the researcher used Cohen's

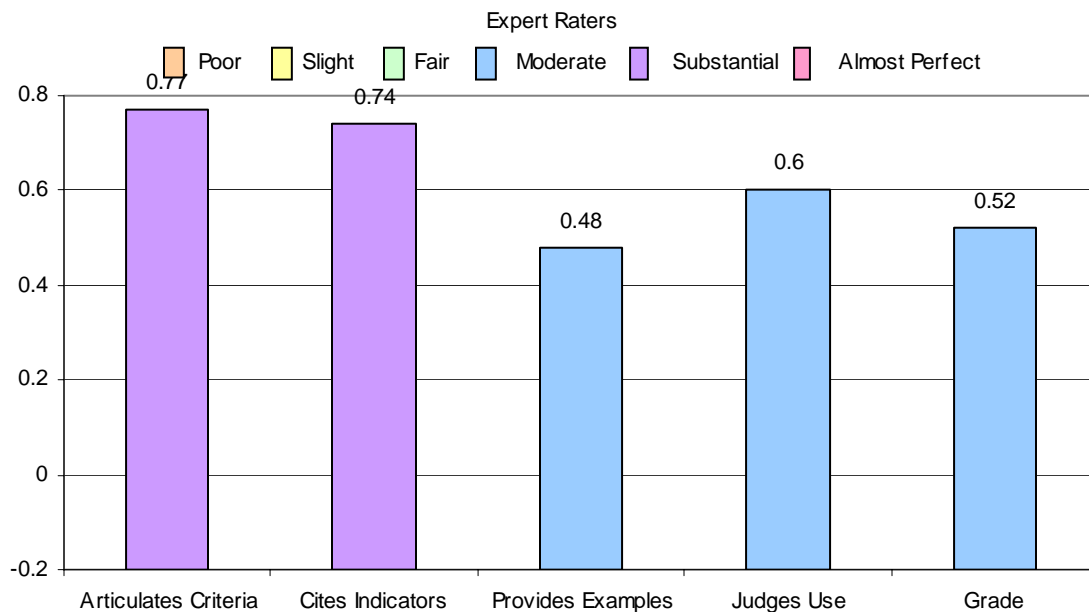


kappa to determine the top 5 raters of the study. This was accomplished by comparing each rater's scores with the "gold standard" set by the researcher (see Chapter 5 for an explanation of this approach) for each of the four criteria included in the study rubric. Then, the kappas for the four rubric criteria were averaged for each rater. Finally, the raters were placed in rank order (see Figure 6.11), and the top five raters were selected as the "expert" raters.

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

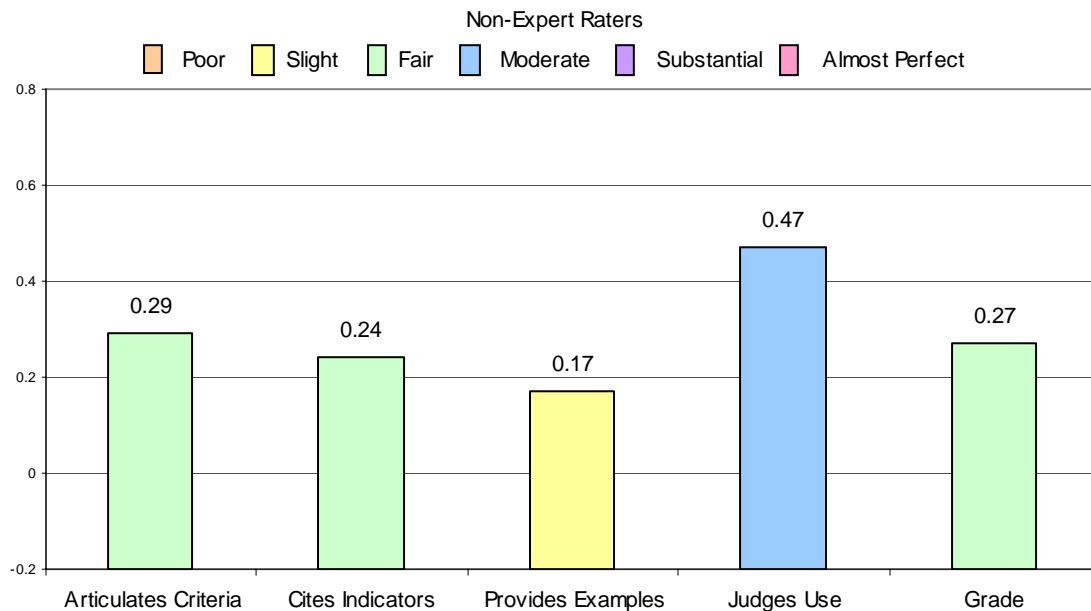
*Figure 6.11. Rank Order of Raters*

The expert raters produced a higher level of reliability than all other rater groups (see Figure 6.12). For two of the four criteria included in the study rubric, the expert rater group achieved substantial levels of agreement. For the first criterion, “Articulates Criteria”, the experts’ scores yielded a kappa of .77 (standard error .03), and the second criterion, “Cites Indicators of Criteria”, the experts’ kappa was .74 (standard error .03). For the other two criteria, the experts showed a moderate level of agreement. For the third criterion, “Links Indicators to Examples from Source”, expert raters’ scores showed kappa of .48 (standard error .03). For the final criterion, “Judges Whether or Not to Use Source”, their kappa was .60 (standard error .03). The expert raters agreed at a moderate level for the overall grade assigned to student responses as well with a kappa of .52 (standard error .03). This expert rater group demonstrated higher reliabilities than any of the pre-established rater groups.



*Figure 6.12. Kappa Statistics for Expert Rat*

In order to compare the proficiency of the expert rater group, the kappa statistics of the non-expert group were also calculated (see Figure 6.13). The reliability of the non-expert group was substantially lower. For two of the rubric criteria, “Articulates Criteria” and “Cites Indicators of Criteria”, non-expert raters showed only fair reliability with kappas of .29 and .24 respectively (standard error .006). For the third criterion, non-experts’ kappa statistics indicated a fair level of agreement ( $k = .17$ , standard error .006). While the kappa for the fourth criterion was moderate, the non-expert raters showed only a fair level of agreement on the overall grade assigned to student responses ( $k = .27$ , standard error .005).



*Figure 6.13. Kappa Statistics for Non-Expert Raters*

The application of t-tests to this data shows that the reliability of all ratings for expert raters was significantly greater than all ratings for non-expert raters. For the first criterion, “Articulates Criteria”, the expert raters achieved a

substantial kappa of .77 while the non-expert raters provided only fair agreement ( $t = 15.7$ ). For the second rubric criterion, "Cites Indicators of Criteria", the kappa for expert raters was .74. This kappa is also significantly different ( $t = 16.3$ ) from the non-expert raters ( $k = .24$ ). Likewise, expert raters produced significantly greater reliability for the third ( $t = 10.1$ ) criterion, "Links Indicators to Examples from Source". For this criterion, the experts showed a kappa of .48 and the non-experts yielded only a .17. A significant difference also appears for the fourth criterion ( $t = 4.2$ ). For "Judges Whether or Not to Use Source," experts showed a kappa of .60 and the non-experts produced a .47. Finally, expert raters yielded moderate levels of agreement ( $k = .52$ ) when assigning overall grades to student responses, a significant difference ( $t = 8.2$ ) from the fair level of agreement provided by non-expert raters ( $k = .27$ ).

It is interesting to note that an analysis of the ten "practice" responses that raters completed before scoring the 75 study responses did not clearly identify these expert raters. Because there were only ten responses scored for practice, there was not enough data to use the kappa statistic effectively. Instead, the practice responses were analyzed using the two-way cross-classification tables generated in SAS when the kappa statistic is run. These cross-classification tables were analyzed to estimate the percent agreement for each rater with the researcher. This method was less precise than the analysis run on the study responses and resulted in a number of "ties". For example, five raters had a 90 percent agreement with the researcher and five had an 88 percent agreement. These ties made it more difficult to determine levels of expertise. However, it

should be noted that three of the five “expert” raters ranked at the top of the percent-agreement ranking. The remaining two ranked slightly lower. This method was able to show that the external librarians had the lowest percent-agreement with the researcher. This anecdotal information is interesting, but statistical significance cannot be demonstrated. Therefore, it seems that these practice responses served as just that—practice.

### **Validity of Rubric Assessment**

In order to demonstrate that an approach to assessment is valid, it must first be shown to be reliable. This study demonstrated that the identification and use of expert raters provides reliable scoring of student responses; therefore, an investigation of validity was merited.

The Cohen's kappa statistic used in this study to explore interrater reliability can also be used as a measure of validity. In fact, it is an accepted practice to compare a group of raters to a “gold standard” to check for validity.<sup>1365</sup> Gwet describes the gold standard as the “correct classification of subjects made by an experienced observer.”<sup>1366</sup> When a gold standard approach is used to test for validity, it is assumed that “the researcher knows the ‘correct classification’ that may be due to an expert judgment.”<sup>1367</sup> Gwet explains, “The question that the researcher wants to answer is whether the...raters agree with the standard.

---

<sup>1365</sup> Gwet, Handbook of Inter-Rater Reliability. 202.

<sup>1366</sup> Gwet, Handbook of Inter-Rater Reliability. 202.

<sup>1367</sup> Gwet, Handbook of Inter-Rater Reliability. 223.

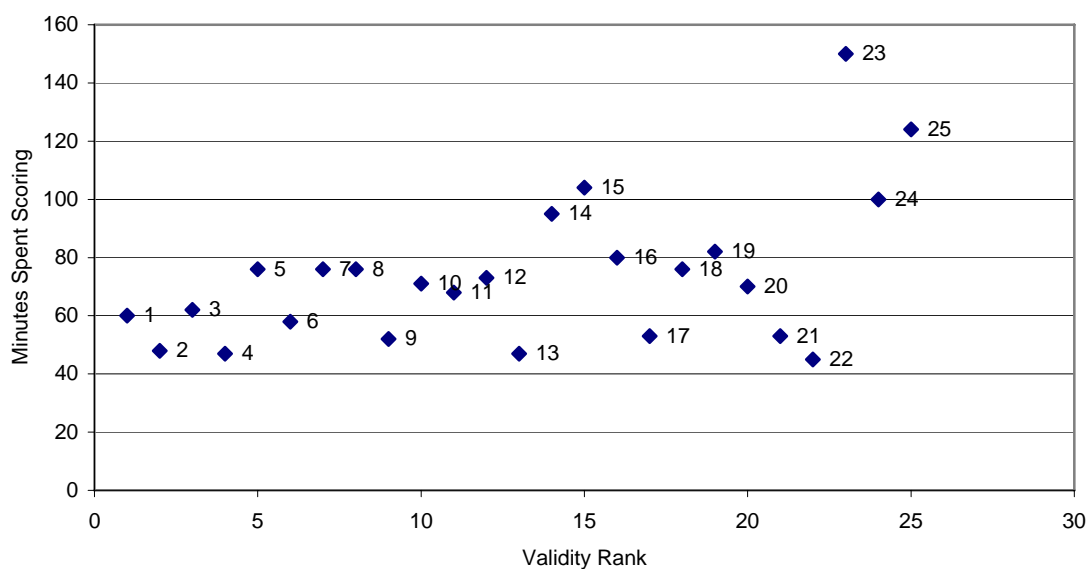
Instead of evaluating the extent of agreement between raters, the researcher wants to know how truthful are the observers' ratings."<sup>1368</sup> In this study, Cohen's kappa was used to compare each rater's scores to the "gold standard" scores (see Figure 6.11). This approach not only identified the expert raters, it also showed the relative validity of the rater groups employed in the study.

As a group, the ENG 101 instructors demonstrated higher validity ranks than other rater groups. In a validity ranking from 1-25 with a rank of 1 as the highest validity rank, the instructors held ranks 2-7. The ENG 101 students also clustered together just below the ENG 101 instructors at ranks 8-13. At the bottom of the rankings were the external reference and instruction librarians, taking ranks 15-25. Unlike the other rater groups the NCSU librarians were scattered throughout the ratings. NCSU librarians were ranked as the 1<sup>st</sup>, 5<sup>th</sup>, 10<sup>th</sup>, 11<sup>th</sup>, and 14<sup>th</sup> raters. These rankings seemed to indicate that (1) ENG 101 instructors produced the most valid scores, followed by ENG 101 students; (2) external librarians produced the least valid scores; and (3) the validity of NCSU librarians' scores varied widely. To describe students' achievement of the learning outcomes addressed in this study, only ratings from the researcher and the 5 most "expert" or valid raters were used.

One additional finding related to the validity ranks of raters was discovered. When the validity rank order was correlated to the minutes each rater spent scoring the 75 student responses using a Pearson correlation statistic,  $r = .50$ . This indicates that a relationship existed between minutes spent

---

<sup>1368</sup> Gwet, Handbook of Inter-Rater Reliability. 223.



*Figure 6.14. Validity Rank and Minutes Spent Scoring*

scoring and scoring accuracy. Figure 6.14 shows the scatterplot of this data.

Each point on the scatterplot is labeled with the validity rank of the scorer.

Because of the small sample size ( $n = 25$ ), it is not possible to draw definitive conclusions based on these results; however, this relationship could be explored in future rubric studies.

### **Student Use of Authority as a Criterion for Website Evaluation**

In order to assess student performance, measures that are both reliable and valid must be employed. In this study, a rubric approach to information literacy skills assessment was found to be reliable and valid when expert raters were used. Therefore, the results described in this section are limited to the assessments of the researcher and the five expert raters identified by the study.

By examining the assessments of student performance performed by both the researcher and the expert raters, much was learned about students' ability to use authority as a criterion to evaluate a website.

This study defined the evaluation of a website using authority as the demonstration of four behaviors. These behaviors are supported by the ACRL Information Literacy Competency Standards for Higher Education, the ACRL Objectives for Information Literacy Instruction: A Model Statement for Academic Librarians, and the LOBO Information Literacy Skills Objectives and Outcomes. These behaviors formed the four criteria of the study rubric and were included as assumptions of the study.

### ***“Articulates Criteria”***

The first behavior students should demonstrate when evaluating a website for authority is the use of criteria terminology. LOBO outcome 3.1.1 states, “The student will articulate established evaluation criteria.” Standard 3.2 of the ACRL document reads, “The information literate student articulates and applies initial criteria for evaluating both the information and its sources.”<sup>1369</sup> Standard 3.2.a. states that students should “evaluate reliability, validity, accuracy, authority, timeliness, and point of view or bias.”<sup>1370</sup> To measure these outcomes, the first criterion on the study rubric classified student responses into three performance descriptions. The first performance level was “Beginning”. Categorized at this

---

<sup>1369</sup> American Library Association, Information Literacy Competency Standards.

<sup>1370</sup> American Library Association, Information Literacy Competency Standards.



level were student responses that did not address authority issues. The second performance description, “Developing”, included student responses that addressed authority issues, but did not use criteria terminology. Finally, student responses that addressed authority issues and used criteria terminology such as author, authority, authorship, or sponsorship were categorized as “Exemplary”.

The distribution of student performance for this behavior is summarized in Figure 6.15. Both the researcher and the expert raters from the study determined that almost 2/3 of students performed at the Developing level and almost 1/3 performed at the Exemplary level. This demonstrates that 2/3 of the students addressed authority issues when evaluating a website, but that only 1/3 used terminology precise criteria terminology such as author, authority, authorship, or sponsorship.

Evaluation Criteria	Beginning		Developing		Exemplary	
	Researcher	Experts	Researcher	Experts	Researcher	Experts
Articulates Criteria	0%	5.9%	68%	64.5%	32%	29.6%

*Figure 6.15. Distribution of Students' Scores for "Articulates Criteria"*

### ***"Cites Indicators of Criteria"***

The second behavior students should demonstrate when evaluating a website for authority is the identification of specific indicators of authority. LOBO outcome 3.1 states, “The student will apply criteria to analyze information, including...authority...to information and its source.” ACRL standard 3.2.c states

that students should apply “evaluative criteria to information and its source.”<sup>1371</sup>

To measure these outcomes, the second criterion of the study rubric classified student responses into three categories. At the Beginning level, students did not address indicators of authority. At the Developing level, students referred vaguely or broadly to authority indicators, but did not cite specific indicators of authority. To be classified at the Exemplary level, students must have cited specific indicators of authority such as domain; server, publisher, host, or the presence of a “~” in the URL; presence of a personal or corporate author name, email, “About Us”, or “Contact Us” links; or author credentials.

Figure 6.16 shows the distribution of student performance for this behavior. Although the researcher and expert raters showed slight differences in the classification at the Beginning and Developing levels of this criterion, both the researcher and the expert raters categorized 3/4 of student responses as

Evaluation Criteria	Beginning		Developing		Exemplary	
	Researcher	Experts	Researcher	Experts	Researcher	Experts
Cites Indicators of Criteria	12%	5.9%	9.3%	19.7%	78.6%	74.4%

*Figure 6.16. Distribution of Students’ Scores for “Cites Indicators of Criteria”*

Exemplary. This indicates that most students cited specific indicators of authority in websites.

---

<sup>1371</sup> Association of College and Research Libraries, Objectives for Information Literacy Instruction.

### ***“Links Indicators to Examples from Source”***

When evaluating a website for authority, students should demonstrate the ability to link indicators of authority to examples from the website they are evaluating. LOBO outcome 3.1.2 and ACRL standard 3.2.a both state that students must “investigate an author’s qualifications and reputation.”<sup>1372</sup> To assess these learning outcomes, the third criterion of the study rubric evaluated students’ ability to find examples of the indicators of authority in the website they were evaluating for use. Student responses that did not address examples of authority indicators from the website under consideration were classified as a Beginning performance. Students who referred vaguely or broadly to examples of authority indicators from the site they were evaluating but did not provide specific examples were categorized as Developing. To be classified as an Exemplary performance, students’ responses must include specific examples of authority indicators located in the site under consideration.

The distribution of student responses across levels of student performance is shown in Figure 6.17. Both the researcher and the expert raters agreed that over 90% of students demonstrated an Exemplary performance for this criterion. Most students located and identified specific examples of authority indicators in the websites they evaluated.

### ***“Judges Whether or Not to Use Source”***

The final behavior students should demonstrate when evaluating a

---

<sup>1372</sup> Association of College and Research Libraries, Objectives for Information Literacy Instruction.

Evaluation Criteria	Beginning		Developing		Exemplary	
	Researcher	Experts	Researcher	Experts	Researcher	Experts
Links Indicators to Examples from Source	5.3%	1.9%	1.3%	6.7%	93.3%	91.5%

*Figure 6.17. Distribution of Students' Scores for "Links Indicators to Examples from Source"*

website for authority is determining whether or not the authority of the site is appropriate for the purpose at hand, usually the completion of an academic paper or project. LOBO outcome 3.2 states that students "will evaluate sources...for use," and LOBO outcome 3.2.2 reads, "The student will indicate whether or not a specific, individual source...is appropriate for the purpose at hand and provide a rationale for that decision based on established evaluation criteria." ACRL standard 3.4.g indicates that students should apply "established evaluation criteria to decide which information sources are most appropriate," distinguish "among various information sources in terms of established evaluation criteria," and describe "why not all information sources are appropriate for all purposes."<sup>1373</sup> The study rubric assessed these outcomes by providing guidelines for the classification of artifacts of student learning. The rubric groups students who did not indicate whether or not the website under consideration was appropriate for the purpose at hand were grouped at the Beginning level of performance. Students who indicated whether or not the site was appropriate to use for the purpose at hand, but did not provide a rationale for that decision that

<sup>1373</sup> Association of College and Research Libraries, Objectives for Information Literacy Instruction.

cited authority issues or indicators were categorized as Developing. The Exemplary classification included students who indicated whether or not the site was appropriate for use and who provided a rationale for that decision based on authority issues or indicators.

The distribution of student responses in Figure 6.18 shows the ratings assigned by the researcher and expert raters in this study. Nearly 1/4 of the student responses fell into the Beginning category because students did not indicate whether or not the website they were evaluating was appropriate for their purposes. Approximately 1 in 5 students judged whether or not the website was appropriate, but did not clearly link that judgment to their evaluation of the authority of the site. Only about 50% of the students observed in this study used an evaluation of authority to determine whether or not a website was appropriate for the purpose at hand.

Evaluation Criteria	Beginning		Developing		Exemplary	
	Researcher	Experts	Researcher	Experts	Researcher	Experts
Judges Whether or Not to Use Source	26.6%	27.7%	22.6%	19.2%	50.6%	53.1%

*Figure 6.18. Distribution of Students' Scores for "Judges Whether or Not to Use the Source"*

### **Summary of Results**

This study sought to answer a number of questions regarding the reliability and validity of a rubric approach to the assessment of information literacy skills. It also attempted to assess an information literacy outcome captured in artifacts

of student learning. These are the answers to the major research questions posed by the study:

Question:

To what degree can different groups of raters provide consistent scoring of artifacts of student learning using a rubric? Can raters provide scores that are consistent with others in their rater group?

Answer:

The answer to this question varies by rater group:

- Within their rater group, NCSU librarians provided moderately reliable scores for the first, second, and fourth criteria on the rubric and the overall grade assigned to the student responses, but they had difficulty coming to consensus on the third criterion of the rubric, “Links Indicators to Examples from Source.”
- Within their rater group, ENG 101 instructors were able to produce moderately to substantially reliable scores in all areas of the rubric and in the total grade assigned to student responses. ENG 101 instructors had the greatest within group reliability of all rater groups studied.
- Within their rater group, ENG 101 students were able to achieve moderate levels of reliability for each criterion included in the rubric, but only achieved a fair level of agreement on the total grade assigned to student responses.

- Within their rater group, internal raters demonstrated a generally moderate level of agreement. Only the third rubric criterion, “Links Indicators to Examples from Source”, produced a fair kappa statistic.
- Within their rater groups, external librarians (instruction and reference) were unable to achieve greater than slight to fair levels of agreement in the first, second, and third criteria of the rubric. The exception was the fourth criterion of the rubric, “Judges Whether or Not to Use Source”. For this criterion, both groups of external librarians achieved a moderate level of agreement. The external instruction librarians demonstrated a fair level of agreement on the total grade assigned to student responses. The external reference librarians demonstrated slight agreement in the same area.
- Overall, the large group including all raters could not achieve desirable levels of agreement.

In summary, the internal rater groups (NCSU librarians, ENG 101 instructors, and ENG 101 students) provided moderately consistent scores with others in their rater groups. In contrast, external librarians could not achieve acceptable levels of agreement.

Question:

Can raters provide scores that are consistent across groups?

Answer:

Several statistically significant differences were discovered by comparing rater groups:

- In nearly all areas, ENG 101 instructors achieved higher levels of reliability than NCSU librarians. The exception to this is the fourth criterion of the study rubric, “Judges Whether or Not to Use Source”. In this area, there was no statistically significant difference between the reliability of scores assigned by the ENG 101 instructors and the NCSU librarians.
- In nearly all areas, the ENG 101 instructors produced higher levels of reliability than the ENG 101 students. However, for the third rubric criterion, “Links Indicators to Examples from Source”, there was no statistically significant difference in the reliability of their scores.
- In most areas, there was no statistically significant difference between the reliabilities of NCSU librarians and ENG 101 students. However, ENG 101 students showed higher reliability for the third criterion of the rubric, “Links Indicators to Examples from Source”. On the other hand, NCSU librarians achieved higher levels of reliability for the fourth criterion on the rubric, “Judges Whether or Not to Use Source”.
- In all areas of the rubric and in the total grade assigned to student responses, internal raters demonstrated higher levels of reliability than external raters.
- In most areas, there was no statistically significant difference between the reliabilities of external instruction librarians and external reference librarians.



- In nearly all areas, NCSU librarians achieved higher levels of reliability than the external librarians. The one exception to this is the third rubric criterion, “Links Indicators to Examples from Source”. In this area, there was no statistically significant difference between the reliability of scores assigned by the NCSU librarians and the external librarians.
- In all areas of the rubric and in the total grade assigned to student responses, expert raters demonstrated higher levels of reliability than non-expert raters.

In summary, ENG 101 instructors produced significantly higher reliabilities than NCSU librarians and ENG 101 students. In fact, few significant differences were discovered between the levels of agreement exhibited by the NCSU librarians and ENG 101 students. However, NCSU librarians produced much higher levels of agreement than the external instruction librarians and external reference librarians. (The two external groups were generally indistinguishable.) Overall, internal raters produced significantly higher levels of agreement. Finally, the raters identified as experts produced the most reliable results of all groups.

Question:

To what degree can raters provide scores that are consistent with scores assigned by the researcher?

Answer:

Some raters, identified through the calculation of the kappa statistic and use of a “gold standard” approach to the examination of validity, provided scores that were quite consistent with the researcher. Indeed, a group of “expert” raters were identified and used to explore the information literacy skills demonstrated by the student responses to the information literacy tutorial, LOBO. This expert group was comprised of two NCSU librarians and three ENG 101 instructors. These raters not only approximated the scores given by the researcher, they also achieved moderate to substantial levels of agreement among their expert rater group. Of the rater groups used at the outset of the study, ENG 101 instructors provided the most valid scores, followed by the ENG 101 students. External librarians demonstrated the lowest levels of validity, and the validity of scores provided by NCSU librarians varied widely.

Question:

To what degree are students able to use authority as a criterion to evaluate a website? Are students able to use precise criteria terminology to address the authority of a website?

Answer:

Approximately 2/3 of the students studied addressed authority issues when evaluating a website, but only about 1/3 used precise criteria terminology such as author, authority, authorship, or sponsorship.

Question:

Are students able to cite specific indicators of authority?

Answer:

Most (3/4) students studied cited specific indicators of authority in websites.

Question:

Are students able to cite examples of authority indicators in the websites they evaluate?

Answer:

Most (approximately 9 out of 10) students studied located and identified specific examples of authority indicators in the websites they evaluated.

Question:

Are students able to decide whether or not a site is appropriate for use and provide a rationale for their decision?

Answer:

Nearly 1/4 of the students studied did not indicate whether or not the website they were evaluating was appropriate for their purposes. Approximately 1 in 5 students stated whether or not the website was appropriate, but did not clearly link that decision to the evaluation of the authority of the site. Only about 1/2 of

the students observed in this study used an evaluation of authority to determine whether or not a website was appropriate for the purpose at hand.

## **CHAPTER 7**

### **DISCUSSION**

This chapter provides a discussion of the results of this study. It is divided into nine sections. First, the purpose of the study and main research questions are recapped. Next, the findings of this study regarding the reliability and validity of a rubric approach to information literacy is summarized. Afterwards, the limitations of the study are detailed. Then the implications of the study are discussed; this section describes the identification of expert raters, the barriers some librarians encounter in becoming expert raters, and the training required to help librarians overcome these barriers. Finally, the value of a rubric approach to information literacy assessment is reaffirmed and recommendations for future research are provided.

#### **Purpose of the Study**

Academic librarians need new approaches to the assessment of information literacy skills. Prior approaches including satisfaction surveys and input/output measures do not provide librarians with enough information about students' achievement of information literacy outcomes. Other approaches including standardized multiple-choice tests and large-scale performance assessments fail to offer librarians the feedback required to improve instruction on a local level. Faced with calls for accountability and armed with the sincere

desire to improve student learning, librarians require new approaches to library instruction assessment. To meet this need, the study was designed to investigate the viability of a rubric approach to information literacy assessment.

To this end, the study documented the use of an information literacy rubric designed to assess student information literacy skills. The analytic rubric used in the study examined students' ability in a specific information literacy skill area—the ability to evaluate the authority of a website. Students exhibited this measurable behavior by responding to open-ended questions in an online tutorial at North Carolina State University. Student responses to the tutorial were collected and assessed by 25 raters in five groups: NCSU librarians, NCSU English 101 instructors, NCSU English 101 students, instruction librarians from other ARL institutions, and reference librarians with some instruction responsibilities from other ARL institutions.

This research design allowed the examination of the consistency with which the rubric scores are assigned by different groups of raters. The design also allowed the researcher to examine the validity of the rubric by comparing the scores given by the raters to the scores given by the researcher. Finally, the study used the rubric assessments of the raters who produced the most reliable and valid scores to assess students' ability to evaluate the authority of a website.

### **Research Questions and Answers**

This study answered the following questions:

1. To what degree can different groups of raters provide consistent scoring of artifacts of student learning using a rubric?
  - a. Can raters provide scores that are consistent with others in their rater group?
  - b. Can raters provide scores that are consistent across groups?
2. To what degree can raters provide scores that are consistent with scores assigned by the researcher?
3. To what degree are students able to use authority as a criterion to evaluate a website?
  - a. Are students able to use precise criteria terminology to address the authority of a website?
  - b. Are students able to cite specific indicators of authority?
  - c. Are students able to cite examples of authority indicators in the websites they evaluate?
  - d. Are students able to decide whether or not a site is appropriate for use and provide a rationale for their decision?

### **Consistent Scoring of Artifacts of Student Learning**

This study demonstrates that multiple raters can use rubrics to produce consistent scoring of artifacts of student learning in the area of information literacy. Because information literacy skills are difficult to articulate, teach, and assess, this is a noteworthy finding! Although educational researchers universally acknowledge the instructional benefits of rubrics, the widespread use

of rubrics to assess information literacy depends on the ability to defend the reliability and validity of the approach, not just in other academic disciplines but also in the field of information literacy instruction. This study is the first to explore the use of rubrics to assess information literacy skills using a large number of raters, and much can be learned from an examination of the results.

While consistent scoring of artifacts of student learning can be achieved using rubrics, different rater groups in this study arrived at varying levels of agreement within their groups. For example, the ENG 101 instructors achieved the greatest levels of agreement within a group, levels that were significantly higher than any other five original rater groups. The ENG 101 instructors were the only one of five original rater groups to attain moderate or substantial levels of agreement across all areas of the study rubric and the grade assigned to each of the student responses. Although definitive reasons for this group's success must await future research, it seems likely that the ENG 101 instructors' familiarity with rubrics used to assess writing may have increased their ability to come to agreement. It is also possible that the ENG 101 instructors, through educational background or personal experience, have learned the values of outcomes-based assessment. As teachers, they are also likely to value the ability to produce consistent scores for complex artifacts of student learning.

ENG 101 students also produced moderate levels of agreement across all areas of the research rubric. This level of consistency might be attributed to previous experiences with rubrics. Although it is probable that the study rubric was students' first experience with a rubric designed to assess information



literacy skills, they may have transferred their earlier rubric experiences to their activities in this study. At the K-12 level, it is common practice to share grading rubrics with students. This self-assessment practice is adhered to for several reasons. Arming students with scoring rubrics allows them to understand their assignments and how they are evaluated. Lazzaro states that rubrics empower students by ensuring that students understand the content of scoring rubrics and deliver their best work.<sup>1374</sup> Herman, Aschbacher, and Winters describe the importance of discussing scoring rubrics with students. They write:

Public discussions help students to internalize the standards and 'rules' they need to become independent learners.... Examples of what constitutes good work engage students in the work itself and in judgments about their work. Public discussions of quality and criteria inform students during the formative period of instruction, not simply at the end of a unit or course when it is too late to make improvements. Furthermore, discussions of criteria also help students see the perspectives of their teachers, their peers, and sometimes even of the experts in the field.<sup>1375</sup>

Shepard asserts that providing students with access to evaluation criteria is required for "basic fairness".<sup>1376</sup> Shepard also believes that students should understand scoring criteria so well that they can evaluate their own work exactly as their teachers would.<sup>1377</sup> When students participate in self-assessment, there are many benefits. Stevens and Levi state that rubrics "encourage students to think critically about their own learning."<sup>1378</sup> According to Shepard, rubrics help

---

<sup>1374</sup> Lazzaro, "Empowering Students with Instructional Rubrics." VI-3:3.

<sup>1375</sup> Herman, Aschbacher and Winters, "Setting Criteria." VI-4:3.

<sup>1376</sup> Shepard, "The Role of Assessment in a Learning Culture." 11-12.

<sup>1377</sup> Shepard, "The Role of Assessment in a Learning Culture." 11.

<sup>1378</sup> Stevens and Levi, Introduction to Rubrics. 21.

students benefit from increased responsibility for learning and a more collaborative relationship between teachers and students.<sup>1379</sup> Not only can students assess themselves using rubrics, they can also work cooperatively with peers to assess each other, as they did in this study. When students evaluate themselves and their peers with rubrics, the process helps them “identify their own learning and development or absence thereof.”<sup>1380</sup>

This study is not the first to explore students’ use of rubrics. In other academic content areas, researchers have investigated the reliability of rubrics in the hands of students. Hafner conducted a study to investigate the use of rubrics by students whom he considered “pedagogically naïve raters.”<sup>1381</sup> Using the instructor’s scores as a comparison to measure validity, he determined that student scores were able to predict the instructor’s scores accurately.<sup>1382</sup> Hafner also found “significant inter-rater reliability” across the study period of 3 years.<sup>1383</sup> He further discovered that the ability to use the rubric was “gender neutral” and had no bearing on students’ overall academic strength in the course.<sup>1384</sup> Hafner concluded that using rubrics for peer assessment “provides an effective teaching and learning strategy.”<sup>1385</sup> This study supports Hafner’s findings and extends the

---

<sup>1379</sup> Shepard, "The Role of Assessment in a Learning Culture." 12.

<sup>1380</sup> Bresciani, Zelná and Anderson, Assessing Student Learning and Development. 31.

<sup>1381</sup> Hafner, "Quantitative Analysis of the Rubric." 1510.

<sup>1382</sup> Hafner, "Quantitative Analysis of the Rubric." 1514, 1520.

<sup>1383</sup> Hafner, "Quantitative Analysis of the Rubric." 1518-1519.

<sup>1384</sup> Hafner, "Quantitative Analysis of the Rubric." 1519.

<sup>1385</sup> Hafner, "Quantitative Analysis of the Rubric." 1526.

research on rubric reliability using student raters into the field of information literacy instruction. Interestingly, one finding of this study does not mesh with Hafner's findings. Although students in this study were able to come to moderate agreement on the four criteria of the rubric, the grades they assigned to overall student responses were less consistent and achieved only a fair level of agreement. This finding—that students may agree on how to use parts of a rubric but not come to adequate levels of consensus on a grade—should be explored in future research.

As a group, NCSU librarians produced moderately consistent scores in most areas of the assessment. Interestingly, there was little significant difference between the levels of reliability achieved by NCSU librarians and ENG 101 students; however, in two areas, students and NCSU librarians differed significantly. For the fourth criterion of the rubric, “Judges Whether or Not to Use the Source”, librarians produced greater levels of agreement than students. For the third criterion of the rubric, “Links Indicators to Examples from Source”, NCSU librarians achieved less consistency than students. The reason that NCSU librarians came only to a fair level of agreement on this criterion is unclear, and should be investigated in future research. Aside from this weak area, the NCSU librarians demonstrated far greater reliabilities within their rater group than the external instruction librarians and external reference librarians.

External librarians (both instruction librarians and reference librarians) could not produce consistent scoring of student responses in this study. It is important to note that there was no statistically significant difference between the

overall performance of external reference librarians and external instruction librarians. These raters' experiences as librarians did not appear to be as important as the fact that they were external to the assessment environment. Taken as a group, external librarians were unable to achieve more than fair levels of agreement on all areas of the assessment, with the exception of the fourth rubric criterion, "Judges Whether or Not to Use Source." (For this fourth criterion, external librarians achieved moderate levels of agreement, but this does not indicate a particular level of expertise in this area. All rater groups were able to achieve a moderate level of agreement for this criterion.) Overall, the external librarians achieved significantly lower levels of reliability than the NCSU librarians and the internal raters as a whole.

In addition to providing important findings regarding the five original rater groups, this study also documents the identification of an expert rater group. The expert rater group achieved higher levels of agreement than any of the five original groups on all areas of the rubric and the grade assigned to student responses. The raters included in the expert group were identified by exploring the levels of agreement between raters and the researcher.

### **Agreement Between Raters and the Researcher**

This study demonstrates that the validity of raters scores can be determined by examining the level of agreement between raters and a "gold standard" set by a researcher. In this study, Cohen's kappa was used to explore the validity of a rubric approach to information literacy assessment.

Using Cohen's kappa to compare raters' scores to a "gold standard" is an accepted practice to check for validity.<sup>1386</sup> Gwet explains that the gold standard is the "correct classification of subjects made by an experienced observer."<sup>1387</sup> Gwet states, "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."<sup>1388</sup> In this study, each rater's scores were compared to the gold standard set by the researcher, and then the raters were ranked according to their agreement with the researcher's scores. This process demonstrated that, as a group, ENG 101 instructors produced the most accurate scores of student responses. The second most accurate group was the ENG 101 students, followed by the external librarians who provided the least accurate scores. Interestingly, the validity of the scores produced by NCSU librarians varied widely. The most accurate scores were provided by one of the NCSU librarians, but another NCSU librarian provided scores that were no more accurate than external librarians. This extreme variation across NCSU librarians demonstrates that training and familiarity with campus contexts and culture is not enough to ensure valid scoring. This finding indicates that, in order to ensure valid scoring of student artifacts of learning, the raters' scores should be tested for validity

---

<sup>1386</sup> Gwet, Handbook of Inter-Rater Reliability. 202.

<sup>1387</sup> Gwet, Handbook of Inter-Rater Reliability. 202.

<sup>1388</sup> Gwet, Handbook of Inter-Rater Reliability. 223.

before a level of expertise is assumed. In this study, only ratings from the five raters with the highest validity ranks were used to estimate student performance.

### **Students' Use of Authority to Evaluate Websites**

Because rubric assessment yields such descriptive data about what students know and can do, this study provides a detailed picture of students' ability to evaluate websites using authority. The study rubric allows raters to assess students' ability to use authority to evaluate a website by breaking the outcome into four criteria: the ability to (1) use authority terminology, (2) cite examples of indicators of the authority, (3) identify those indicators in an example website, and (4) make a reasoned decision about the use of the example website. This study found that about 2/3 of the students addressed authority issues when evaluating a website, but only about 1/3 used precise terms like "author", "authority", "authorship", or "sponsorship" when doing so. It also found that most students (about 3/4) cited specific indicators of authority such as: domain, presence of an author's name, "About Us" links, or author credentials. Nearly all students (about 9 out of 10) located and cited these indicators in the website they were considering for an academic assignment. Perhaps the most interesting finding was that students had far less success deciding whether or not a website was appropriate for their assignment and providing a rationale for that decision. Nearly 1/4 of the students did not indicate whether or not the site they were evaluating was appropriate for their purposes. About 1 out of 5 students stated whether or not the website was appropriate, but did not clearly link that

decision to the authority of the site. Only about 1/2 of the students observed in this study used an evaluation of authority to determine whether or not a website was appropriate for the purpose at hand.

These results are significant for numerous reasons. First, they provide important data about the information behavior of students. They clarify both the strengths and weaknesses of students' abilities. Second, by making students' achievement of learning outcomes clear to librarians and instructors, the study rubric allows educators to celebrate students' successes and diagnose and plan for problem areas. Based on the findings of this rubric assessment, NCSU librarians and ENG 101 instructors can "close the loop" by using assessment data to improve learning experiences. For example, librarians and instructors might plan changes to the LOBO tutorial, improved integration of the tutorial in ENG 101 classes, or creation of follow up lessons that focus on applying evaluative criteria to decision making. Third, this study provides a foundation for future assessment projects that can diagnose students' difficulty in making decisions based on what they know about the value of information sources.

### **Limitations of the Study**

It is important to recognize that aspects of environment and design limit this or any study. Therefore, the conclusions made from this study are not necessarily generalizable across other populations.

A delimitation of this study is that the participants include English 101 students, English 101 instructors, NCSU librarians, and ten librarians from five

other university systems. Students and instructors in other classes at NCSU and at other universities have not been included. Librarians at other universities, with the exception of those mentioned above, have not been included. Therefore, the findings are not necessarily generalizable to students and instructors in other areas at NCSU, nor are they necessarily generalizable to librarians at other universities.

A limitation of this study is that students, having just received instruction about using the criterion of authority to evaluate a website, may exhibit better than normal behavior due to temporary heightened awareness.

### **Implications of the Study**

Limitations aside, this study has important implications and can be used to improve student performance and establish best practices in information literacy instruction and assessment. Prior to this study, no studies existed in the library literature in which a large number of raters used rubrics to score a large number of artifacts of student learning. Similarly, there were no studies in the library literature that analyzed the interrater reliability of multiple raters among and/or between groups. Also, until this study, there were no studies that examined the validity of a rubric used to assess information literacy skills. As the first study of its kind in the field of information literacy assessment, the research offers significant findings in a number of areas. Beyond its contributions to library literature, this study also adds a new step to the methodology of rater training and suggests new practices for the statistical analysis of interrater reliability.



### ***Not All Raters Are Experts***

One of the most important findings of this study is that not all raters are “expert” raters. While this might seem an obvious observation, in practice many rubrics are used by raters that have not been identified as expert in the use of assessment tools. Often, expert ability in the application of rubrics is assumed based on an individual’s educational background, experience, or position in an educational institution. In fact, the divisions among rater groups used in this study were predicated on education, experience, and position. Librarians were separated from instructors based on the achievement of a library science degree. External reference librarians and external instruction librarians were divided based on experience. Students were separated from educators based on position in the institution. Although interesting and meaningful differences were demonstrated among these groups, the group of “expert” raters crossed divisions of education, experience, and position. Thus, the belief that rubrics can be used reliably and validly by raters who have a certain position in the institution, a particular degree, or a specific type of experience is a dangerous assumption. If raters are selected because of their position, education, or experience rather than demonstrated ability to provide reliable and valid scores, students may be scored inconsistently, inaccurately, and unfairly. This study supports the idea that only expert raters should make instructional decisions that impact learners.

### ***A Method for Identifying Expert Raters***

Since not all raters are expert raters, how can expert raters be identified? This study documents the use of statistics to identify expert raters among a pool of raters, an approach that eschews potentially faulty assumptions of expertise based on education, experience, or position. In this study, the “gold standard” approach to examining validity was successfully used to select raters who could provide the most valid ratings of student work. This approach uses Cohen’s kappa to compare raters, not against other raters, but against a “gold standard” of validity. Those raters who most closely match the standard are declared to be “expert” raters. The ratings of experts can be used to produce a valid assessment of student learning. In this study, the expert rater group not only achieved valid assessments, they also boasted the highest reliability levels of any rater group. The success of the gold standard approach in producing meaningful assessments of student information literacy skills implies that this approach to the selection of raters is a useful practice to follow in future rubric research, especially in the area of information literacy assessment.

### ***Characteristics of Expert Raters***

While this study offers a model for identifying expert raters, it is also worth considering the factors that make an “expert” rater an expert. Why do some raters achieve expert status while others do not?

Some possible answers to this question emerge both from the literature and from rater comments made throughout this study. First, the literature reports

that raters may have diverse outlooks, perspectives, and experiences that need to be taken into account.<sup>1389</sup> Tired raters may not score students evenly over an assessment period.<sup>1390</sup> A rater's mood can affect the scores they assign.<sup>1391</sup> Prior knowledge of a student's ability or attitude can also impact rater's responses to student work.<sup>1392</sup> These factors impact the intrarater reliability of an assessment. Moskal and Leydens suggest that "well-designed scoring rubrics respond to the concern of intrarater reliability by establishing the scoring criteria in advance."<sup>1393</sup> On the other hand, even rubrics that are well designed can be used inconsistently by raters.

Some raters might be naturally proficient scorers. Wolfe, Kao and Ranney describe proficient scorers. They note that the most proficient scorers tend to focus on the general features of a student's product or performance and "adopt values espoused by the scoring rubric" more so than less proficient scorers.<sup>1394</sup> Throughout the scoring process, the rater should revisit the established criteria in order to ensure that consistency is maintained."<sup>1395</sup> However, less proficient scorers tend to interrupt their observation of a student's product or performance

---

<sup>1389</sup> Colton, Gao, Harris, Kolen, Martinovich-Barhite, Wang and Welch, Reliability Issues with Performance Assessments. 9.

<sup>1390</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1391</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1392</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1393</sup> Barbara M. Moskal and Jon A. Leydens, "Scoring Rubric Development: Validity and Reliability," Practical Assessment Research and Evaluation 7.10 (2000).

<sup>1394</sup> Wolfe, Kao and Ranney, "Cognitive Differences."

<sup>1395</sup> Moskal, "Scoring Rubrics: What, When, and How?."

to monitor how well it satisfies the rubric, rather than focusing on the product or performance and then reviewing it against the rubric.<sup>1396</sup>

Training can also impact the proficiency of scorers. Certainly the process of “norming” or “calibrating” raters improves proficiency, and this is born out by the significant differences between internal and external raters who participated in this study. In addition to the lack of training, there may be other barriers that keep raters from becoming expert raters.

In this study, raters who completed the open-ended comment sheet after scoring student responses alluded to some of these roadblocks. Their comments acknowledged the presence in this study of six barriers that may explain why some raters could not attain expert status.

### ***Barrier 1: Difficulty Understanding Outcomes-Based Assessment***

While outcomes-based assessment is moving forward in other parts of institutions of higher education, it remains a challenge in academic libraries. Iannuzzi reports, “We have yet to see widespread implementation of outcomes assessment methodologies in terms of student learning in our academic libraries.”<sup>1397</sup> Libraries continue to struggle with assessment of outcomes approaches to instruction.<sup>1398</sup> Iannuzzi suggests a few explanations for the ongoing challenge of assessing student achievement of information literacy

---

<sup>1396</sup> Wolfe, Kao and Ranney, "Cognitive Differences."

<sup>1397</sup> Iannuzzi, "We Are Teaching, but Are They Learning." 304.

<sup>1398</sup> Iannuzzi, "We Are Teaching, but Are They Learning." 304.

outcomes. She admits that assessment is “difficult and potentially frightening.”<sup>1399</sup> She voices concerns that assessment is too challenging for libraries to complete alone and explains that librarians may be afraid that they will be held solely accountable for any disappointing results.<sup>1400</sup>

One of the main difficulties librarians encounter when they attempt to use outcomes-based assessment tools is that they do not understand the fundamental concepts of outcomes-based education. In fact, some librarians struggle with the most basic concept of the approach: measurable student learning outcomes.

Learning outcomes are “the essential and enduring knowledge, abilities (skills), and attitudes (values, dispositions) that constitute the integrated learning needed by a graduate of a course or program.”<sup>1401</sup> These outcomes are not about “what you are going to do to the student, but rather what you want to student to know or do as a result of an initiative, course or activity.”<sup>1402</sup> Outcomes must be measurable.<sup>1403</sup> In other words, educators must be able to identify or observe student behaviors that show whether or not students know and can do what the program set out to teach them.<sup>1404</sup> To be measurable, they

---

<sup>1399</sup> Iannuzzi, "We Are Teaching, but Are They Learning." 304.

<sup>1400</sup> Iannuzzi, "We Are Teaching, but Are They Learning." 304.

<sup>1401</sup> Battersby, "So, What's a Learning Outcome Anyway?."

<sup>1402</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 11.

<sup>1403</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 11.

<sup>1404</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 11.

usually include active verbs that can be identified or observed by assessors.<sup>1405</sup>

Dugan and Hernon define outcomes in a library context as: “observed, reported, or otherwise quantified changes in attitudes or skills of students on an individual basis because of contact with library services, programs, or instruction.”<sup>1406</sup>

Lichtenstein reports that librarians are increasingly asked to join other educators in justifying their programs using learning outcomes.<sup>1407</sup> Smith 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.”<sup>1408</sup> However, Lichtenstein laments that library instruction outcomes are often unstated or too vague.<sup>1409</sup>

The comments made by librarian raters in this study may serve to explain this conflict between the need for clear measurable library instruction outcomes and the difficulty librarians have in creating and working with outcomes. For instance, a number of librarian raters in this study 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 raters felt that using measurable outcomes to assess student learning focused too much on specific skills and that an outcomes-based approach was too much

---

<sup>1405</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 11.

<sup>1406</sup> Dugan and Hernon, "Outcomes Assessment." 379.

<sup>1407</sup> Lichtenstein, "Informed Instruction." 28.

<sup>1408</sup> Smith, New Roles and Responsibilities for the University Library.

<sup>1409</sup> Lichtenstein, "Informed Instruction." 28.

“science” and not enough “art”. One rater wrote, “While the rubric measures the presence of concepts...it doesn’t check to see if students understand [the] issues.” Another rater 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 meaningful progress can be made in the assessment of information literacy skills.

### ***Barrier 2: Tension Between Analytic and Holistic Approaches***

Outcomes-based assessment of student learning is essentially an analytic approach to assessment because it assumes that student learning can be articulated in outcomes and that those outcomes can be systematically assessed. However, in this study, some librarian raters had difficulty with an analytic approach to the measurement of student learning. One rater commented that using the rubric “was really simple. But I worried that I was being too simplistic...and not rating [student work] holistically.” Another wrote, “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 show a lack of comfort with analytic approaches to the assessment of student learning.

One possible compromise for librarians might be the adoption of more holistic, rather than analytic, rubrics. A holistic rubric “requires the teacher to score the overall process or product as a whole, without judging the component

parts separately.”<sup>1410</sup> It gives one score for a whole product or performance based on an overall impression.<sup>1411</sup> Holistic rubrics provide an “overall, single judgment of quality.”<sup>1412</sup> Popham writes, “A holistic scoring strategy signifies that the scorer must attend to how well a student’s response satisfies all the evaluative criteria in the interest of forming a general, overall evaluation of the response based on all criteria considered in concert.”<sup>1413</sup>

Holistic rubrics have both advantages and disadvantages. Arter and McTighe recommend holistic rubrics for simple products or performances, particularly ones with only one important criterion to assess.<sup>1414</sup> Moskal points out that holistic rubrics are sometimes preferred when the criteria that mark a product or performance are difficult to distinguish or when they overlap.<sup>1415</sup> Arter and McTighe also note that holistic rubrics are useful for “getting a quick snapshot of overall quality or achievement.”<sup>1416</sup> Mertler suggests that holistic rubrics are often used when the focus is on a summative assessment, since they provide only limited feedback.<sup>1417</sup> This points to one main disadvantage of holistic rubrics. They provide “no detailed analysis of the strengths and

---

<sup>1410</sup> Nitko, Educational Assessment of Students. 226.

<sup>1411</sup> Arter and McTighe, Scoring Rubrics. 18.

<sup>1412</sup> Arter and McTighe, Scoring Rubrics. 18.

<sup>1413</sup> Popham, Test Better, Teach Better. 96.

<sup>1414</sup> Arter and McTighe, Scoring Rubrics. 21.

<sup>1415</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1416</sup> Arter and McTighe, Scoring Rubrics. 21.

<sup>1417</sup> Mertler, "Designing Scoring Rubrics."



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."<sup>1418</sup>

The rubric used in this study was an analytic rubric. 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."<sup>1419</sup> They allow for separate evaluations of each factor along a different descriptive scale.<sup>1420</sup> Then, assessors sum the individual scores to form a total score.<sup>1421</sup> Popham writes, "An analytic approach to scoring requires a scorer to make a criterion-by-criterion judgment for each of the evaluative criteria, and then amalgamate those per-criterion ratings into a final score (this is often done via a set of predetermined, possibly numerical, rules)."<sup>1422</sup> Because of this part-to-whole approach, analytic rubrics offer a multidimensional assessment approach.<sup>1423</sup>

Like holistic rubrics, there are both advantages and disadvantages to using an analytic rubric approach to assessment. According to Mertler, analytic rubrics are preferred when a focused response to stakeholders is required.<sup>1424</sup>

---

<sup>1418</sup> Arter and McTighe, Scoring Rubrics. 21.

<sup>1419</sup> Arter and McTighe, Scoring Rubrics. 18.

<sup>1420</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1421</sup> Nitko, Educational Assessment of Students. 266.

<sup>1422</sup> Popham, Test Better, Teach Better. 96.

<sup>1423</sup> Mertler, "Designing Scoring Rubrics."

<sup>1424</sup> Mertler, "Designing Scoring Rubrics."

Analytic rubrics are better suited for “judging complex performances (e.g. research process) involving several significant dimensions.... By breaking such performances down into traits, raters (including students and teachers) can more readily grasp the essential components of quality.”<sup>1425</sup> Mertler states that the advantages of analytic rubrics are “substantial,” also citing the feedback they offer to students that simply “does not happen” when holistic rubrics are used.<sup>1426</sup> He notes that the feedback is detailed enough to create a “profile” of student strengths and weaknesses.<sup>1427</sup> Arter and McTighe suggest that such rich feedback can be used to target future instruction to specific areas of need.<sup>1428</sup> They write, “From an instructional perspective, analytical trait rubrics help students come to better understand the nature of quality work since they identify the important dimensions of a product or performance.”<sup>1429</sup> Popham agrees that analytic rubrics are more diagnostic and notes that they allow an assessment of student performance per-criterion.<sup>1430</sup> All of these advantages come at a cost. Analytical rubrics take more time to create and to use. Arter and McTighe write, “After all, you have more to discern.”<sup>1431</sup>

---

<sup>1425</sup> Arter and McTighe, Scoring Rubrics. 22.

<sup>1426</sup> Mertler, "Designing Scoring Rubrics."

<sup>1427</sup> Mertler, "Designing Scoring Rubrics."

<sup>1428</sup> Arter and McTighe, Scoring Rubrics. 22.

<sup>1429</sup> Arter and McTighe, Scoring Rubrics. 22.

<sup>1430</sup> Popham, Test Better, Teach Better. 97.

<sup>1431</sup> Arter and McTighe, Scoring Rubrics. 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,<sup>1432</sup> 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."<sup>1433</sup> 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 specific improvements in the tutorial or plan for other methods to address deficiencies in student learning. Ultimately, the decision to use an analytic or holistic rubric must depend on the assessment goals of stakeholders.

### ***Barrier 3: Failure to Comprehend the Rubric***

A third issue that may prevent raters from becoming "expert" scores is simple: a failure to understand some aspect of the rubric. Training raters to use a rubric can significantly decrease the likelihood of this barrier to expert use, but sometimes the problem persists.

Nearly all the raters used in this study stated that they understood the words used in the rubric. Four of the 25 commented that they were confused about the differences between "issues" and "indicators", but the vast majority had no difficulty with the language used in the rubric. All the raters except one noted

---

<sup>1432</sup> Mertler, "Designing Scoring Rubrics."

<sup>1433</sup> Mertler, "Designing Scoring Rubrics."

that the rubric was visually appealing. (The one exception requested that any new versions of the rubric have more cell padding so that the grid lines were placed further from the text.) Nearly all of the raters stated that the rubric was free of cultural, ethnic, and gender stereotypes and biases—problems that could certainly lead to rater confusion. The one rater who expressed caution in this area noted that “any emphasis on authority—i.e. authority is good—is a cultural bias.” However, since the ability to evaluate information sources for authority is an ACRL standard, it is likely that most raters would not find the emphasis on authority problematic.

One comprehension problem of greater importance surfaced during this study. A few external librarian raters voiced difficulty in deciding how many times a student response should show achievement of an outcome to be scored at a particular performance level. One of the external raters solved this problem by asserting her personal understanding of the rubric. This librarian wrote, “I decided to use literally examples, indicators to mean that students needed to provide more than one.” Another external rater commented, “The student might cite one example...but not...enough for me to consider it exemplary.” A third rater was bothered by the lack of credit awarded for quantity in the rubric. She wrote, “I personally would prefer a more complex rubric (that would allow scoring for the number of indicators a student cites, for example). The way it’s set up, I was giving the same score to a student who cited every indicator and a student who only cited one.” These concerns about the quantity rather than quality of student demonstrations of learning reveal a lack of experience or knowledge

about rubric design. Callison notes that rubrics that emphasize quantity rather than quality are problematic.<sup>1434</sup> He suggests that rubrics should not judge student performance by how many times a student does something, but on how well he or she does it.<sup>1435</sup> The lack of rubric experience or knowledge that causes this type of comprehension failure might be easily addressed through training.

#### ***Barrier 4: Disagreement with Assumptions of the Rubric***

Two assumptions are fundamental to the rubric used in this study. The first is that information literacy skills can be expressed as student learning outcomes, described in sufficient detail to be captured in artifacts of student learning, and recognized and scored by raters. Second, it is assumed that the ability to use the criterion of authority to evaluate of a website 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 ACRL documents, the Information Literacy Competency Standards for Higher Education and the Objectives for Information Literacy Instruction: A Model Statement for Academic Librarians, that form the basis for the criteria included in the study rubric.

---

<sup>1434</sup> Callison, "Rubrics." 36.

<sup>1435</sup> Callison, "Rubrics." 36.

Perhaps the biggest hurdle for some raters in this study was that they did not agree with some of the standards on which the rubric was based and, as a result, had a difficult time adopting the assumptions of the rubric. The standard that posed the most difficulty for raters was ACRL standard 3.2 which reads, “The information literate student articulates and applies initial criteria for evaluating both the information and its sources.”<sup>1436</sup> To measure the first part of this standard, the articulation of evaluation criteria, the first criterion on the study rubric classified student responses into three performance descriptions. The first performance level was “Beginning”. Student responses that did not address authority issues were categorized at this level. The second performance description, “Developing”, included student responses that addressed authority issues, but did not use criteria terminology. Finally, student responses that addressed authority issues and used criteria terminology such as author, authority, authorship, or sponsorship were categorized as “Exemplary”. Several raters in this study argued that students should not have to articulate the criteria they were looking for. They suggested that the application of the criteria, which is assessed in other parts of the study rubric, was the true test of student learning. Some raters felt that assessing students’ ability to use criteria terminology resulted in “looking for specific terms and missing the overall point of the responses.” One rater 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 rater’s comments seem

---

<sup>1436</sup> American Library Association, Information Literacy Competency Standards.

to indicate that her disagreement with one part of the rubric impacted her ability to adopt the values of the rubric in its entirety.

Two approaches might help eliminate this barrier for raters. First, conflicts with the assumptions of a rubric might be avoided if raters 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. Callison writes, "Rubrics are texts that are visible signs of agreed-upon values. They cannot contain all the nuances of the evaluation community's values, but they do contain the central expressions of those values."<sup>1437</sup>

Including stakeholders in the rubric development process offers a number of advantages. Students can benefit directly from being involved in the development of rubric assessments. Including students in the process of rubric development makes them full participants in the assessment process. Shepard believes that students should understand scoring criteria so well that they can evaluate their own work exactly as their teachers would.<sup>1438</sup> Frederikson and Collin argue that rubrics should make students metacognitively aware of what makes good work good work.<sup>1439</sup> By truly understanding the criteria with which their work will be assessed, students should "perform better than learners who do not have this knowledge."<sup>1440</sup> Including students in the development of

---

<sup>1437</sup> Callison, "Rubrics." 36.

<sup>1438</sup> Shepard, "The Role of Assessment in a Learning Culture." 11.

<sup>1439</sup> Shepard, "The Role of Assessment in a Learning Culture." 11.

<sup>1440</sup> Lazzaro, "Empowering Students with Instructional Rubrics." VI-3:3.

assessment helps “build...ownership of the evaluation, makes it clear that judgments need not be arbitrary, and makes it possible to hold students to higher standards because the criteria are clear and reasonable.”<sup>1441</sup>

Students are not the only stakeholders who could benefit from participation in the rubric development process. Several authors cite collaboration between librarians and instructors as a crucial part of successful information literacy assessment. Rockman states,

Within the higher education environment, it is important for students to be able to build upon a foundation of information literacy skills and abilities by being able to transfer their knowledge from course to course. Responsibility for helping them reach this goal and for assessing their progress is best accomplished through faculty-librarian partnerships.<sup>1442</sup>

Bernier speaks as a librarian stating, “We discovered that creating rubrics with teachers increases our effectiveness. Our voices are heard when we work with teachers to design and teach standards-based lessons. We use rubrics and criteria lists to improve the library research process.”<sup>1443</sup> Carter states that assessment “requires collaboration with faculty in other departments”<sup>1444</sup> and Samson posits that the entire library instruction program “is a collaborative process, and its assessment must address the concerns of all participants.”<sup>1445</sup>

---

<sup>1441</sup> Wiggins, "Creating Tests Worth Taking." V-6:6.

<sup>1442</sup> Rockman, "Integrating Information Literacy." 612.

<sup>1443</sup> Bernier, "Making Yourself Indispensable." 24.

<sup>1444</sup> Carter, "'Doing the Best You Can with What You Have'." 41.

<sup>1445</sup> Samson, "What and When Do They Know? Web-Based Assessment."



Knight summarizes, "Ultimately, the library must attempt to integrate faculty members in our assessment projects."<sup>1446</sup>

Experts from across the academy support such collaborative approaches to assessment. One of the Nine Principles of Good Practice for Assessing Student Learning supports such collaborations across the university. Principle 6 reads, "Assessment fosters wider improvement when representatives from across the educational community are involved. Student learning is a campus-wide responsibility, and assessment is a way of enacting that responsibility. Thus while assessment may start small, the aim over time is to involve people from across the educational community."<sup>1447</sup> Hoyler writes, "Assessment...requires us to work together, and to do unfamiliar things like setting common goals and standards, devising methods of assessment, interpreting the results, and using them to improve and coordinate our teaching."<sup>1448</sup>

Using rubrics based on agreed upon values increases the likelihood of consistent scoring. Callison states, "if applied correctly, rubrics come close to assuring that inadequate, satisfactory, and excellent mean the same thing on the same skill set from one group of students to a similar group regardless of who makes the evaluation."<sup>1449</sup> Including raters in rubric development may significantly improve the ability of experts to espouse the assumptions of rubric.

---

<sup>1446</sup> Knight, "The Role of Assessment in Library User Education."

<sup>1447</sup> American Association of Higher Education, Nine Principles.

<sup>1448</sup> Hoyler, "The Road Not Taken."

<sup>1449</sup> Callison, "Rubrics." 35.

Another possibility is that this barrier originates not with the raters, nor with the rubric used in the study. Instead, the ACRL standards used to develop this rubric might play a role. Because the raters who disagreed with the assumptions of the rubric were in fact disagreeing with the standards themselves, one of the implications of the study might be that there is insufficient agreement among academic librarians on the ACRL standards. In this study, raters' concerns about requiring students to articulate evaluation criteria indicate that the ACRL standards may require revision before they can be used as a foundation for outcomes-based assessment of information literacy skills. Revisions might begin with a reduction of duplication and redundancy, a reconsideration of the behaviors students should demonstrate to indicate achievement of the outcomes, and discussions of what achievement of each outcome "looks like" so that academic librarians can come to agreement on what constitutes a beginning, developing, or exemplary student performance.

#### ***Barrier 5: Difficulties with Artifacts of Student Learning***

The artifacts of student learning present a fifth barrier that may stand in the way of raters producing expert levels of agreement and accuracy. In this study, several raters commented on the difficulty of interpreting student answers that were cryptic, incomplete, vague, or incorrect. One external rater 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." One external rater 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 raters were stymied by incorrect or partially incorrect answers. One rater 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 student artifacts can affect the ability of raters to produce consistent and accurate assessments.

### ***Barrier 6: Difficulties Understanding Campus Context and Culture***

In higher education, external raters are often included in assessment projects and may include "internship advisors, alumni, professionals within the community, or faculty from nearby institutions."<sup>1450</sup> By employing external raters to participate in an assessment of student learning, educators can obtain feedback about the assessment tools they employ or about the overall assessment process. Including external raters has several advantages. For example, external raters can increase impartiality and provide third-party objectivity.<sup>1451</sup> Their feedback can be used for student evaluation, and the use of

---

<sup>1450</sup> Bernier, "Making Yourself Indispensable." 126.

<sup>1451</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 79-80.

external raters may have the added benefit of stimulating collaborative efforts between other departments or institutions.<sup>1452</sup>

External review can also result in a few disadvantages. For instance, there is a risk of a mismatch between the rater's expertise or expectations and the local assessment purpose.<sup>1453</sup> Prus and Johnson also warn that external review can be expensive and time-consuming if individualized evaluations are required or the program being assessed is large.<sup>1454</sup> In order to reduce these disadvantages, Prus and Johnson suggest sharing the philosophy of the program, student learning outcomes, and criteria for assessment beforehand.<sup>1455</sup>

In this study, external raters played an important part of the research design, which approximated the realities of assessment in academic libraries. In academic libraries, assessments are either created on campus where training is available, or assessments are imported from a separate institution and only written materials are available for consultation. To replicate conditions encountered in real life, the internal raters in this study participated in a training session. In contrast, the external raters were provided with a substantial amount of background material, directions, and examples to familiarize them with the campus culture and study context—content that might be found in an article or written report.

---

<sup>1452</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 79-80.

<sup>1453</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 80.

<sup>1454</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 80.

<sup>1455</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 80.

In this study, external raters were unable to demonstrate acceptable levels of consistency in the scoring of student responses, and they achieved the lowest levels of validity when compared to a gold standard. Such low levels of reliability and validity could be due to lack of training or a lack of familiarity with NCSU conventions and culture. Because this research was designed to approximate real life rather than contrived lab-like conditions, it is not possible to determine whether the difficulties of external librarians can be attributed to the absence of training or their lack of first-hand knowledge regarding NCSU. What is clear is that the typical model for exporting an assessment that works on one campus to another campus, via professional listservs or journals, may not produce reliable and valid assessment results. This is an important finding with implications for academic librarians' professional practice.

### **The Need for Training**

Although instructors and students exhibit a level of proficiency in the use of rubrics, this study demonstrates that librarians are less proficient. Fortunately, this study also indicates that internal librarians can be trained to become expert raters. Rubrics that are developed in a campus context and culture can deliver detailed information about students' achievement of information literacy outcomes. If librarians hope to harness the power of rubric assessment, they must undergo training that gives them a foundation of educational concepts and principles and, in so doing, helps them overcome the barriers to proficient teaching and assessment.

This study revealed several barriers to expert rating that should be addressed by training. First, librarians need to learn the values and principles of outcomes-based assessment. Many of the principles of outcomes-based assessment addressed in Chapter 3 could be included in such training. Librarians can also address this barrier by forming partnerships with faculty teaching and learning support programs, assessment units, and schools of education on their campuses.

Training for librarians should include content that reviews the theories that underlie rubric assessment, including the advantages and disadvantages of various rubric models, such as analytic and holistic rubrics. Training should include acknowledgement of the structural issues that can limit the reliability, validity, and overall instructional usefulness of rubrics. Some rubrics are not well written;<sup>1456</sup> some use wording that is too general or too specific;<sup>1457</sup> some are too long;<sup>1458</sup> some include inconsistencies;<sup>1459</sup> and some emphasize quantity rather than quality.<sup>1460</sup> Training should familiarize librarians with the content and structure of other information literacy rubrics and their strengths and weaknesses. By examining other information literacy rubrics, librarians can avoid past mistakes and improve the quality of information literacy rubrics over time.

---

<sup>1456</sup> Popham, Test Better, Teach Better. 95.

<sup>1457</sup> Tierney and Simon, "What's Still Wrong with Rubrics."

<sup>1458</sup> Popham, Test Better, Teach Better. 98.

<sup>1459</sup> Tierney and Simon, "What's Still Wrong with Rubrics."

<sup>1460</sup> Callison, "Rubrics." 36.

Maki's model of calibration is a useful training tool that can be used to address issues that affect raters' comprehension of the rubric.<sup>1461</sup> To Maki's model, perhaps one step should be added. This study contributes an additional step in the process by discovering that modeling the application of the rubric to student responses using "think aloud" techniques is an effective initial step. In this study, the researcher talked through her use of the study rubric to assess five anchor responses. This prompted discussion among raters and brought norming issues out early in the calibration process. Maki's model also calls for two to three rounds of rater scoring and discussion.<sup>1462</sup> In this study, two rounds were used before raters felt confident about scoring student responses independently. In hindsight, it is possible that a third round may have been useful in coming to agreement on the handling of deficiencies in the formats of some student responses. It also may be that multiple calibration sessions should be required for librarians who have spent little or no time using rubrics in the past.

Training should also address disagreement with assumptions of the rubric. In this study, raters voiced concerns about one of the ACRL standards included in the rubric, and it is possible that these concerns made it difficult for some raters to adopt the values of the rubric in its entirety. Training can address this problem in two ways. Training can begin earlier in the rubric development process if stakeholders and raters are included in the creation of the rubric from

---

<sup>1461</sup> Maki, Assessing for Learning. 126.

<sup>1462</sup> Maki, Assessing for Learning. 127.

the beginning. Also, when information literacy rubrics are predicated on ACRL standards, training should include consensus building on which standards to include and how to operationalize them. That is, in order to come to agreement on the assessment of student achievement, librarians must agree on what behavior looks like when a student achieves the learning outcomes being assessed.

Despite the best efforts of educators, sometimes the artifacts of student learning that are analyzed in assessment processes are less complete or less sophisticated than desired. Certainly, one of the purposes of assessment processes is to identify these weaknesses in student responses to learning activities so that they can be addressed. However, deficiencies in artifacts of student learning can make it difficult for raters to apply rubrics, as a few raters in this study noted. It is important that training address the difficulties that raters are likely to encounter. Careful selection of anchor responses is an important part of addressing the ways in which raters should score particularly poor or incomplete student responses. If training does not address this problems that surface in artifacts of student learning, raters are likely to devise their own ways of scoring such papers and decrease the reliability of scoring overall.

Training about educational concepts related to instruction and assessment will allow librarians to become full partners in the teaching and learning processes of educational institutions. According to Owusu-Ansah, “Librarians...must accept formally their teaching role and engage actively in it, not sporadically, but as a generally accepted mandate of the profession and of



the academic library in academe.”<sup>1463</sup> This is true, not just of teaching in general, but also of the part of teaching that is assessment. It is imperative that librarians seek training in educational theories and principles and become actively engaged in the assessment of student learning.

### **The Value of Rubrics**

If so much training is required to help librarians use rubrics to produce reliable and valid assessments, skeptics might ask if the benefits of rubrics are worth the time and energy training requires. Certainly, there are costs associated with training, and the major cost is time. Prus and Johnson acknowledge the potential cost of time, but state that the advantages of rubric assessment outweigh this cost.<sup>1464</sup> The raters that participated in this research study agree. Their comments reveal that they believe that rubrics have great instructional value. All the internal raters stated that they could envision using the rubric to improve information literacy instruction, and all but one of the external raters agreed.

Research confirms the instructional benefits of rubrics. Popham calls rubrics “instructional illuminators” and emphasizes that “appropriately designed rubrics can make an enormous contribution to instructional quality.”<sup>1465</sup> According to Lichtenstein, rubrics make learning easier and more effective

---

<sup>1463</sup> Owusu-Ansah, "Information Literacy and Higher Education." 12.

<sup>1464</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 25.

<sup>1465</sup> Popham, "What's Wrong--and What's Right--with Rubrics." 75.

because, “people usually gain more from a lesson when they are told what it is they are supposed to learn.”<sup>1466</sup> Resnick writes, “As educators, it is our job to help students understand how to construct their own learning and thereby continue to be life-long learners. The use of rubrics as a teaching and learning tool can plan an integral part in attaining this goal.”<sup>1467</sup>

The specific instructional benefits of rubrics include direct benefits to students, the benefits of stating agreed upon values, and the benefit of descriptive result data. Rubrics benefit students by allowing them to understand the expectations of their instructors. Rubrics provide direct feedback to students about what they have learned and what they have yet to learn. Hafner describes the importance of the top level of a rubric. He writes, “Importantly, the top level of the rubric communicates what exemplary work should look like and, as such, involves the student in constructive learning and self-evaluation.”<sup>1468</sup> Finally, rubrics help students understand why a particular grade is given to a student’s product or performance.<sup>1469</sup>

Rubrics benefit all instructional stakeholders by providing an opportunity to discuss and determine agreed upon values of student learning. Callison writes, “Rubrics are texts that are visible signs of agreed-upon values. They cannot contain all the nuances of the evaluation community’s values, but they do contain

---

<sup>1466</sup> Lichtenstein, “Informed Instruction.” 28.

<sup>1467</sup> Resnick, Education and Learning to Think.

<sup>1468</sup> Hafner, “Quantitative Analysis of the Rubric.” 1510.

<sup>1469</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.

the central expressions of those values.”<sup>1470</sup> Stevens and Levi list the facilitation of communication with others as a key reason to use rubrics.<sup>1471</sup> Using rubrics based on agreed upon values has practical advantages for consistent scoring. First, “if applied correctly, rubrics come close to assuring that inadequate, satisfactory, and excellent mean the same thing on the same skill set from one group of students to a similar group regardless of who makes the evaluation.”<sup>1472</sup>

Rubric assessments provide detailed descriptions of what students know and can do. Rubrics provide “evaluators and those whose work is being evaluated with rich and detailed descriptions of what is being learned and what is not.”<sup>1473</sup> This descriptive data can be used to document how educators or program administrators improve instruction.<sup>1474</sup> Because rubrics are easy to use and easy to explain, the data they generate is not only rich and descriptive, but also easy to understand, defend, and convey.<sup>1475</sup>

Rubrics are also useful for assessing standards-based instruction programs and student learning over time. According to Bresciani, Zelna and Anderson, rubrics teach students “the standards of the discipline or the standards of the cocurricular learning and development experience.”<sup>1476</sup> They help

---

<sup>1470</sup> Callison, "Rubrics." 36.

<sup>1471</sup> Stevens and Levi, Introduction to Rubrics. 23.

<sup>1472</sup> Callison, "Rubrics." 35.

<sup>1473</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 30.

<sup>1474</sup> Bernier, "Making Yourself Indispensable." 25.

<sup>1475</sup> Andrade, "Using Rubrics." 14.

<sup>1476</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.

examine the extent to which students achieve standards.<sup>1477</sup> Rubrics can measure achievement of standards from a formative perspective or an “ecological” one, revealing how students “apply knowledge in authentic situations.”<sup>1478</sup> This is significant because accreditors seek “multiple dimensions of student performance,” not just a summative grade.<sup>1479</sup>

Additionally, rubrics can be used to assess student learning over time. Educators can use the same or similar rubrics over time to assess student progress toward learning goals.<sup>1480</sup> Rubrics can be used to assess individual students, all students in a program, or students across multiple programs.<sup>1481</sup> This is especially important for educators who need to capture the learning of a student population that flows from program to program.<sup>1482</sup>

Although this study indicates that librarians require substantial training in order to use rubrics consistently and accurately, the benefits associated with rubric assessment far outweigh the time spent in training. Prus and Johnson write: “As in virtually all other domains of human assessment, there is a consistently inverse correlation between the quality of measurement methods and their expediency; the best methods usually take longer.”<sup>1483</sup> Librarians who

---

<sup>1477</sup> Moskal, "Scoring Rubrics: What, When, and How?."

<sup>1478</sup> Pausch and Popp, "Assessment of Information Literacy."

<sup>1479</sup> Council for Higher Education Accreditation, "Student Learning Outcomes Workshop." 2.

<sup>1480</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.

<sup>1481</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 31.

<sup>1482</sup> Bresciani, Zelna and Anderson, Assessing Student Learning and Development. 35.

<sup>1483</sup> Prus and Johnson, "A Critical Review of Student Assessment Options." 25.

seek to develop meaningful measures of student achievement should avoid faster approaches that fail to offer the instructional benefits of rubric assessment. Instead, librarians should learn the educational concepts, principles, and theories that will allow them to use rubrics to their fullest potential. To do otherwise is to discard a viable approach to assessment that offers a wealth of benefits to students and librarians alike.

### **Recommendations for Future Research**

Because this study is the first of its kind in the area of information literacy instruction, the findings described in this study await testing and confirmation by future researchers. In addition to replicating this research in other environments, several other areas of research should be explored.

The data generated in this study might be used for future investigations. For instance, the data might be inverted so that researchers could attempt to determine what types of student responses might be more likely to be scored consistently or inconsistently. Future research could also attempt to determine what characteristics made the group of ENG 101 instructors the most successful raters in this study. Finally, the tendency of ENG 101 students to agree on how to score student responses on parts of the rubric but not the final grade assigned to the response bears future scrutiny.

Other findings from this study may merit further investigation. For example, a relationship appears to exist between the amount of time that raters took to score student responses and the accuracy of their scores. This study

suggests that accurate scorers actually take less time to score student responses. However, the total number of raters was insufficient to test the significance of that tendency. Second, all raters in the study appeared to produce greater levels of agreement as they scored. Thus, the consistency with which raters scored the second 25 student responses of the study was higher than the first 25 responses, and the last 25 responses were scored more consistently than the second 25. In a study that included more raters, researchers could determine if or why some rater groups followed this pattern and others did not.

Other areas pertaining to the rubric assessment of information literacy skills should be pursued as well. Further research should attempt to discover the fewest number of student artifacts scores needed to identify expert raters. The attributes of expert raters should also be examined so that they may be identified in a rater pool more quickly. Future researchers might also consider exploring the effects of different types and levels of rater training and attempt to discover what special training might be required to transform librarians into expert users of information literacy rubrics. For instance, future researchers could attempt to evaluate the reliability and validity of scores provided by internal librarians who participated in training and those who did not. Future research might also attempt to explore the effects of different types and levels of rater training in external librarian populations.

Ultimately, this study is intended to inspire researchers to investigate the viability of a rubric approach to the assessment of information literacy skills.

Future researchers might pursue the rubric assessment of other information literacy outcomes, or examine existing information literacy rubrics using the methodology employed in this study. Whatever future research follows, it is hoped that this study will provide a foundation for those who pursue the promise of rubrics for the assessment of information literacy skills.

## APPENDIX A

### LOBO Information Literacy Objectives and Outcomes

#### Objective 1

**The information literate student will determine the nature and extent of an information need.**

##### Outcome 1.1

The student will develop a realistic overall plan and timeline to acquire the needed information. (ACRL 1.3.c)

- 1.1.1 The student will describe the stages of the research process. (ACRL 2.2.a) I, NDM
- 1.1.2 The student will search for, gather, and synthesize information based on an informal, flexible plan. (ACRL 1.3.c) I, NDM
- 1.1.3 The student will act appropriately to obtain information within the time frame required. (ACRL 1.3.c) I, NDM

##### Outcome 1.2

The student will define and articulate the need for information. (ACRL 1.1)

- 1.2.1 The student will identify an initial research topic. (ACRL 1.1.d) I, P
- 1.2.2 The student will narrow or broaden the scope or direction of the topic to achieve a manageable focus. (ACRL 1.1.d) I, P
- 1.2.3 The student will list key concepts and terms describing the facets of the research topic that may be useful in locating information. (ACRL 1.1.e) I, P
- 1.2.4 The student will narrow, broaden, or refine key concepts and terms describing the research topic. (ACRL 1.1.e) I, P
- 1.2.5 The student will demonstrate an understanding of how the desired end product will play a role in determining the need for information. (ACRL 1.1.d, 1.4.b) I, P
- 1.2.6 The student will describe how the intended audience influences information choices. (ACRL 1.4.b) I, P
- 1.2.7 The student will explore general information sources to increase familiarity with the topic. (ACRL 1.1.c) I, TBD

##### Outcome 1.3



The student will identify a variety of potential sources for needed information. (ACRL 1.2)

1.3.1 The student will identify various formats in which the information is available. (ACRL 1.2.c) I, P

1.3.2 The student will identify the value and differences (e.g., purpose, audience) of potential resources in a variety of formats. (ACRL 1.2.c, 1.2.d) I, P

## **Objective 2**

**The information literate student will access needed information effectively and efficiently.**

### **Outcome 2.1**

The student will construct and implement effectively-designed search strategies. (ACRL 2.2)

2.1.1 The student will identify related terms and synonyms for the research topic. (ACRL 2.2.b) I, P

2.1.2 The student will identify phrases to use as search terms for the research topic. (ACRL 2.2.b) I, P

2.1.3 The student will identify alternative endings, abbreviations, and multiple spellings of search terms for the research topic. I, P

2.1.4 The student will construct search statements using Boolean operators. (ACRL 2.2.d) I, P

2.1.5 The student will identify search terms to truncate, if appropriate. (ACRL 2.2.d) I, P

### **Outcome 2.2**

The student will select the most appropriate retrieval method or system for accessing needed information. (ACRL 2.1)

2.2.1 The student will use different research sources (e.g., search engines, databases, catalogs) to find different types of information (e.g., web sites, articles, books). (ACRL 2.3.a) I, P

2.2.2 The student will describe the differences between article databases and library catalogs and/or search engines. (ACRL 2.1.c) I, P

2.2.3 The student will distinguish among article databases, identifying what types (e.g., general, subject-specific) or subject coverage is most appropriate for a research topic. (ACRL 2.1.c) I, P

- 2.2.4 The student will identify differences between basic and advanced interfaces in search engines, when more than one interface is available. (ACRL 2.2.e) I, P

Outcome 2.3

The student will retrieve information online or in person using a variety of methods. (ACRL 2.3)

- 2.3.1 The student will use the LC call number system locate resources within the library. (ACRL 2.3.b) I, P
- 2.3.2 The student will determine whether or not a cited item is available immediately. (ACRL 1.3.a, 2.5.c) I, P

**Objective 3**

**The information literate student will evaluate information critically.**

Outcome 3.1

The student will apply criteria to analyze information, including authority, content, purpose, timeliness, and point of view or bias, to information and its source (ACRL 3.2.a, 3.2.c)

- 3.1.1 The student will articulate established evaluation criteria. (ACRL 3.2.a) I, P
- 3.1.2 The student will investigate an author's qualifications and reputation. (ACRL 3.2.a) I, P
- 3.1.3 The student will investigate a publisher or issuing agency's qualifications and reputation. (ACRL 3.2.a, 3.4.e) I, P
- 3.1.4 The student will describe the content of an information source. I, P
- 3.1.5 The student will describe the purpose for which information was created. (ACRL 3.2.d) I, P
- 3.1.6 The student will identify where to look for a source's publication date and, if possible, determine when the information was published. (ACRL 3.2.a) I, P
- 3.1.7 The student will articulate the importance of timeliness or currency and/or describe the impact of the age of a source or the qualities characteristic of the time in which it was created. (ACRL 3.2.a, 3.2.d) I, P
- 3.1.8 The student will recognize prejudice, deception, or manipulation. (ACRL 3.2.c) I, P

- 3.1.9 The student will articulate the impact of an author's, sponsor's, and/or publisher's point of view. (ACRL 3.2.c) I, P
- 3.1.10 The student will describe how cultural, geographic, or other contexts within which the information was created may bias information. (ACRL 3.2.d) I, P
- 3.1.11 The student will recognize the presence of one-sided views, opinions, emotional triggers, stereotypes, etc. (ACRL 3.2.c) I, P
- 3.1.12 The student will consider the impact of his/her own biases on his/her interpretation of information. I, P
- 3.1.13 The student will investigate a source's point of view or bias through comparison with other sources, including links, citations found in the source, or other similar sources. (ACRL 3.2.a, 3.7.c) TBD
- 3.1.14 The student will distinguish scholarly from popular sources. I, P

#### Outcome 3.2

The student will evaluate sources (e.g., article, web site, book, journal, database, catalog) for use. (ACRL 3.4.g)

- 3.2.1 The student will determine whether or not various information sources (e.g. web sites, popular magazines, scholarly journals, books) are appropriate for the purpose at hand, based on established evaluation criteria (see LOBO 3.1), and provide a rationale for that decision. (ACRL 3.4.g, 2.2.a) I, P
- 3.2.2 The student will indicate whether or not a specific, individual source (e.g., a particular web site, article, book) is appropriate for the purpose at hand and provide a rationale for that decision based on established evaluation criteria (see LOBO 3.1). (ACRL 3.2.c) I, P

#### **Objective 4**

**The information literate student will use information effectively to accomplish a specific purpose.**

ENG 101 instructors are responsible for setting and achieving outcomes related to this objective.

#### **Objective 5**

**The information literate student will use information ethically and legally.**

#### Outcome 5.1

The student will integrate their research into learning products without plagiarizing.

5.1.1 The student will articulate the differences among the acceptable methods for integrating research (e.g., quoting, summarizing, paraphrasing). I, P

5.1.2 The student will identify when to use acceptable methods for integrating research. I, P

#### Outcome 5.2

The student will acknowledge the use of information sources through documentation styles. (ACRL 5.3.a)

5.2.1 The student will locate information about documentation styles. (ACRL 5.3.a) I, P

5.2.2 The student will select an appropriate or assigned documentation style among different styles. (ACRL 5.3.a) I, P

5.2.3 The student will identify citation elements for information sources in different formats (e.g., book, scholarly journal article, web site, interview). (ACRL 5.3.a) I, P

5.2.4 The student will follow documentation style guidelines correctly and consistently. (ACRL 5.3.a) I, P

---

I = Introduced  
P = Practiced

NDM = Not Directly Measured  
TBD = To Be Developed

## APPENDIX B

### Study Rubric

Evaluation Criteria	Beginning	Developing	Exemplary	Student Learning Outcomes
<b>Articulates Criteria</b>	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.	<b>LOBO 3.1.1</b> The student will articulate established evaluation criteria. (ACRL 3.2.a)
<b>Cites Indicators of Criteria</b>	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.	<b>LOBO 3.1</b> The student will apply criteria to analyze information, including...authority...to information and its source (ACRL 3.2.a, 3.2.c)
<b>Links Indicators to Examples from Source</b>	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.	<b>LOBO 3.1</b> The student will apply criteria to analyze information, including...authority...to information and its source (ACRL 3.2.a, 3.2.c) <b>LOBO 3.1.2</b> The student will investigate an author's qualifications and reputation. (ACRL 3.2.a)
<b>Judges Whether or Not To Use Source</b>	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.	<b>LOBO 3.2</b> The student will evaluate sources (e.g., article, web site, book, journal, database, catalog) for use. (ACRL 3.4.g) <b>LOBO 3.2.2</b> The student will indicate whether or not a specific, individual source (e.g., a particular web site, article, book) is appropriate for the purpose at hand and provide a rationale for that decision based on established evaluation criteria (see LOBO 3.1). (ACRL 3.2.c)

## APPENDIX C

### Sample Score Sheet

Student Response # \_\_\_\_\_ Position # \_\_\_\_\_ of 75 Rater # \_\_\_\_\_

[INSERT STUDENT RESPONSE HERE]

Evaluation Criteria	Beginning	Developing	Exemplary
<b>Articulates Criteria</b>	0 - Student does not address authority issues. <div style="text-align: right;">○</div>	1 - Student addresses authority issues, but does not use criteria terminology. <div style="text-align: right;">○</div>	2 - Student addresses authority issues and uses criteria terminology such as: author, authority, authorship, or sponsorship. <div style="text-align: right;">○</div>
<b>Cites Indicators of Criteria</b>	0 - Student does not address authority indicators. <div style="text-align: right;">○</div>	1 - Student refers vaguely or broadly to authority indicators, but does not cite specific indicators. <div style="text-align: right;">○</div>	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. <div style="text-align: right;">○</div>
<b>Links Indicators to Examples from Source</b>	0 - Student does not address examples of authority indicators from the site. <div style="text-align: right;">○</div>	1 - Student refers vaguely or broadly to examples of authority indicators from the site under consideration, but does not cite specific examples. <div style="text-align: right;">○</div>	2 - Student cites specific examples of authority indicators from the site under consideration. <div style="text-align: right;">○</div>
<b>Judges Whether or Not To Use Source</b>	0 - Student does not indicate whether or not the site is appropriate to use for the purpose at hand. <div style="text-align: right;">○</div>	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. <div style="text-align: right;">○</div>	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. <div style="text-align: right;">○</div>

RESEARCHER USE ONLY: Total Score \_\_\_\_/8

**Respond to the following prompts in the space below, using complete sentences:**

- Identify the "domain type" of the site you're evaluating and explain why that is acceptable or unacceptable for your needs.
- Identify the "publisher" or host of the site and tell what you know (or can find out) about it.
- State whether or not the site is a personal site and explain why that is acceptable or unacceptable for your needs.
- State who (name the person or institution) created the site and tell what you know (or can find out) about the creator.
- Look for the author's credentials on the site. List his/her credentials and draw conclusions based on those credentials. If there are no credentials listed, tell what conclusions you can draw from their absence.
- Using what you know about the AUTHORITY of this web site, explain why it is or is not appropriate to use for your paper/project.

## APPENDIX D

### Internal Rater Training Power Point

**Welcome!**

Grab a name tag!  
Find your seat!

The Assessment of Student  
Information Literacy Skills:  
A Rubric Approach

September 24, 2005

## The Purpose of the Study

## The Need

- Librarians and professors need to be able to show that students have information literacy skills.
- What is information literacy?

## The Information Literate Student

- **Determines** the nature and extent of the **information needed**.
- **Accesses** needed information effectively and efficiently.
- **Evaluates** information and its sources critically.
- **Uses** information effectively to accomplish a specific purpose.
- Accesses and uses information **ethically** and **legally**.



## The Need, cont'd.

- We need a new approach to assessing students' information literacy skills.
- Traditional approaches have flaws.
  - Standardized Tests
  - Some Performance Assessments
- There's another option:  
Rubric assessment has potential and merits investigation.

## The Purpose of the Study

To investigate the viability of a  
rubric approach  
to information literacy assessment.

## A Brief Introduction to Rubrics

(and how rubrics can be used with LOBO)

## What is a rubric?

- A **chart** that describes student work or behavior according to “**outcomes**”.

Examples:

“The student will articulate established evaluation criteria.”

“The student will evaluate sources for use.”

- Rubrics outline the essential parts of student work or behavior (known as “**criteria**”).
- Rubrics describe what good and poor “**performances**” look like.

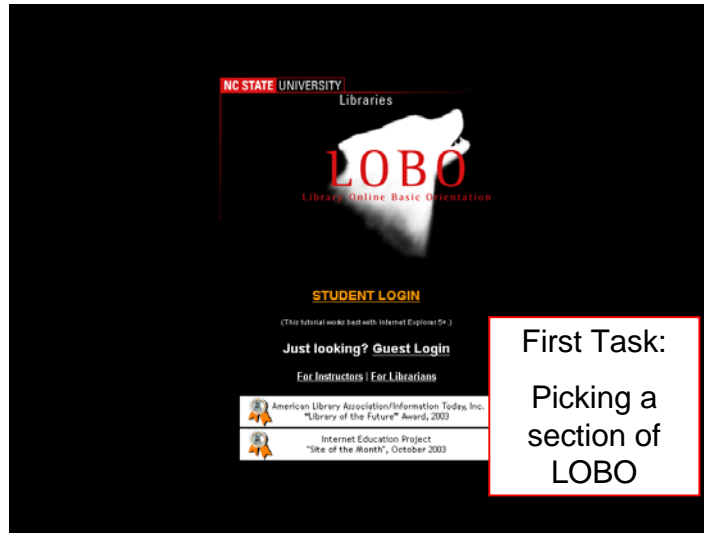
	Poor	So-So	Good
Criteria 1	Description of doing a poor job on Criteria 1	Description of doing a so-so job on Criteria 1	Description of doing a good job on Criteria 1
Criteria 2	Description of doing a poor job on Criteria 2	Description of doing a so-so job on Criteria 2	Description of doing a good job on Criteria 2

## Starting Small...

1 Rubric for  
1 Part of  
LOBO

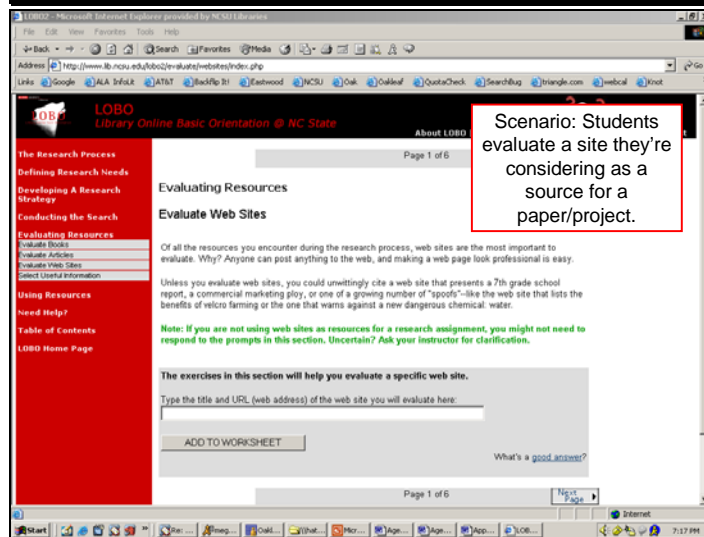


I hope I can be used  
for other types of  
assignments too!

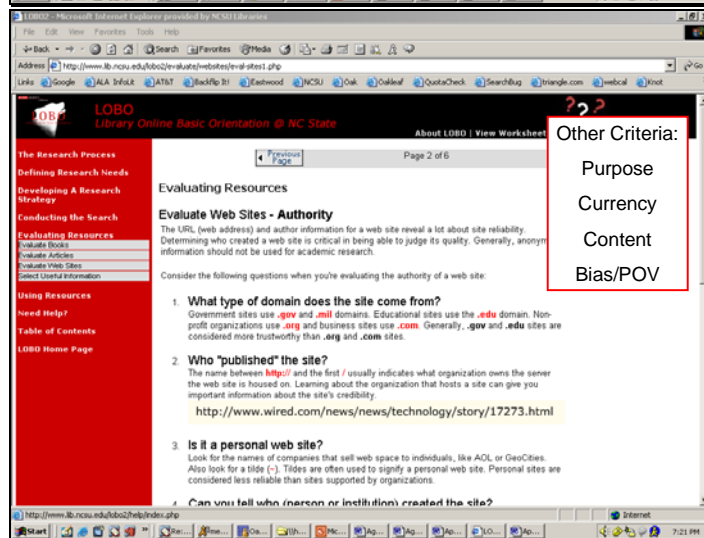


First Task:

Picking a section of LOBO



Scenario: Students evaluate a site they're considering as a source for a paper/project.



Other Criteria:

Purpose  
Currency  
Content  
Bias/POV

**Evaluating Web Sites - Authority**

The LOBO team selected the following questions to help you evaluate a website's authority. Information about a website's authority is found in a variety of places on the website. Consider the following questions when you're evaluating the authority of a web site.

- 1. What type of domain does the site come from?**  
Look for the domain type (e.g., .edu, .gov, .org, .com, .net, .mil). Each domain type has its own characteristics. For example, .edu domains are usually associated with educational institutions, .gov domains with government, and .com domains with commercial entities.
- 2. Who "published" the site?**  
The name of the publisher or host of the site usually indicates what organization owns the site. The site is hosted on a server, and the organization that owns the site can give you more information about the site's reliability.  
<http://www.wired.com/wired/story/>
- 3. Is it a personal web site?**  
Look for the name of an individual that will appear in the site. If it is a personal site, look for a name in the site. If it is a personal site, look for a name in the site. If it is a personal site, look for a name in the site.
- 4. Can you tell who created or published the site?**  
Look for the name of the person or institution that created the site. If it is a personal site, look for a name in the site. If it is a personal site, look for a name in the site.
- 5. Are the author's credentials listed on the site?**  
If you can find the author's name on a site, look for the author's name in the site. If you can find the author's name on a site, look for the author's name in the site.

**Respond to the following prompts in the space below, using complete sentences:**

- Identify the "domain type" of the site you're evaluating and explain why that is acceptable or unacceptable for your needs.
- Identify the "publisher" or host of the site and tell what you know (or can find out) about it.
- State whether or not the site is a personal site and explain why that is acceptable or unacceptable for your needs.
- State who (name the person or institution) created the site and tell what you know (or can find out) about the creator.
- Look for the author's credentials on the site. List his/her credentials and draw conclusions based on those credentials. If there are no credentials listed, tell what conclusions you can draw from their absence.
- Using what you know about the AUTHORITY of this web site, explain why it is or is not appropriate to use for your paper/project.

ADD TO WORKSHEET

How might an instructor score your answer?

**Questions**

**Respond to the following prompts in the space below, using complete sentences:**

- Identify the "domain type" of the site you're evaluating and explain why that is acceptable or unacceptable for your needs.
- Identify the "publisher" or host of the site and tell what you know (or can find out) about it.
- State whether or not the site is a personal site and explain why that is acceptable or unacceptable for your needs.
- State who (name the person or institution) created the site and tell what you know (or can find out) about the creator.
- Look for the author's credentials on the site. List his/her credentials and draw conclusions based on those credentials. If there are no credentials listed, tell what conclusions you can draw from their absence.
- Using what you know about the AUTHORITY of this web site, explain why it is or is not appropriate to use for your paper/project.

ADD TO WORKSHEET

How might an instructor score your answer?

## What's our rubric based on?

- The "outcomes" we designed LOBO to teach.
- The outcomes for LOBO came from the standards produced by the Association of College and Research Libraries.
- Ideally, using accepted, national standards makes our rubric more generalizable to multiple situations, not just LOBO.

5.3.2 Authority		Outcomes		
Evaluation Criteria	Beginning	Developing	Exemplary	Student Learning Outcomes
Articulates Criteria	0 - Student does not address authority issues.	1 - Student addresses authority issues, but does not use criteria terminology such as: author, authority, authorship, or sponsorship.	2 - Student addresses authority issues and uses criteria terminology such as: author, authority, authorship, or sponsorship.	LOBO 3.1.1 The student will articulate established evaluation criteria. (ACRL 3.2.a)
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 such as: domain, server, or ~ in URL; presence of personal or corporate author name, email, "About Us" or "Contact Us" links; or author credentials.	2 - Student cites specific authority indicators such as: domain, server, or ~ in URL; presence of personal or corporate author name, email, "About Us" or "Contact Us" links; or author credentials.	LOBO 3.1 The student will apply criteria to analyze information, including...to information and its source (ACRL 3.2.a, 3.2.c)
Links Indicators to Examples from Source	0 - Student does not cite 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.	LOBO 3.1 The student will apply criteria to analyze information, including...to information and its source (ACRL 3.2.a, 3.2.c) LOBO 3.1.2 The student will investigate an author's qualifications and reputation. (ACRL 3.2.a)
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.	LOBO 3.2 The student will evaluate sources (e.g., article, web site, book, journal, database, catalog) for use. (ACRL 3.4.g) LOBO 3.2.2 The student will indicate whether or not a specific, individual source (e.g., a particular web site, article, book) is appropriate for the purpose at hand and provides a rationale for that decision based on established evaluation criteria (see LOBO 3.1). (ACRL 3.2.c)

## What criteria did we use?

- 4 parts of web site evaluation:
  - Using Criteria Terminology
  - Citing Criteria Indicators
  - Citing Examples of Indicators from the Site
  - Judging Whether or Not to Use the Site

5.3.2 Authority		Criteria		
Evaluation Criteria	Beginning	Developing	Exemplary	Student Learning Outcomes
Articulates Criteria	0 - Student does not address authority issues.	1 - Student addresses authority issues, but does not use criteria terminology such as: author, authority, authorship, or sponsorship.	2 - Student addresses authority issues and uses criteria terminology such as: author, authority, authorship, or sponsorship.	LOBO 3.1.1 The student will articulate established evaluation criteria. (ACRL 3.2.a)
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 such as: domain, server, or ~ in URL; presence of personal or corporate author name, email, "About Us" or "Contact Us" links; or author credentials.	2 - Student cites specific authority indicators such as: domain, server, or ~ in URL; presence of personal or corporate author name, email, "About Us" or "Contact Us" links; or author credentials.	LOBO 3.1 The student will apply criteria to analyze information, including...to information and its source (ACRL 3.2.a, 3.2.c)
Links Indicators to Examples from Source	0 - Student does not cite 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.	LOBO 3.1 The student will apply criteria to analyze information, including...to information and its source (ACRL 3.2.a, 3.2.c) LOBO 3.1.2 The student will investigate an author's qualifications and reputation. (ACRL 3.2.a)
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.	LOBO 3.2 The student will evaluate sources (e.g., article, web site, book, journal, database, catalog) for use. (ACRL 3.4.g) LOBO 3.2.2 The student will indicate whether or not a specific, individual source (e.g., a particular web site, article, book) is appropriate for the purpose at hand and provides a rationale for that decision based on established evaluation criteria (see LOBO 3.1). (ACRL 3.2.c)

## What levels of performance did we include?

- Exemplary → Meets outcome completely.  
What a “good” answer looks like.
- Developing → Shows progress toward meeting outcome, but does not meet it completely.  
What a “medium” answer looks like.
- Beginning → Does not meet outcome.  
What a “poor” answer looks like.

### 5.3.2 Authority

Evaluation Criteria	Beginning	Developing	Exemplary	Student Learning Outcomes
<b>Articulates Criteria</b>	0 - Student does not address authority issues.	1 - Student addresses authority issues, but does not use criteria terminology such as: author, authority, authorship, or sponsorship.	2 - Student addresses authority issues and uses criteria terminology such as: author, authority, authorship, or sponsorship.	<b>LOBO 3.1.1</b> The student will articulate established evaluation criteria. (ACRL 3.2.a)
<b>Cites Indicators of Criteria</b>	0 - Student does not address authority indicators.	1 - Student refers vaguely or broadly to authority indicators such as: domain, server, or ~ in URL; presence of personal or corporate author name, email, "About Us" or "Contact Us" links; or author credentials.	2 - Student cites specific authority indicators such as: domain, server, or ~ in URL; presence of personal or corporate author name, email, "About Us" or "Contact Us" links; or author credentials.	<b>LOBO 3.1</b> The student will apply criteria to analyze information, including...to information and its source (ACRL 3.2.a, 3.2.c)
<b>Links Indicators to Examples from Source</b>	0 - Student does not cite 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.	<b>LOBO 3.1</b> The student will apply criteria to analyze information, including...to information and its source (ACRL 3.2.a, 3.2.c) <b>LOBO 3.1.2</b> The student will investigate an author's qualifications and reputation. (ACRL 3.2.a)
<b>Judges Whether or Not To Use Source</b>	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.	<b>LOBO 3.2</b> The student will evaluate sources (e.g., article, web site, book, journal, database, catalog) for use. (ACRL 3.4.g) <b>LOBO 3.2.2</b> The student will indicate whether or not a specific, individual source (e.g., a particular web site, article, book) is appropriate for the purpose at hand and provide a rationale for that decision based on established evaluation criteria (see LOBO 3.1). (ACRL 3.2.c)

Sharing the Concept with Students

Link to Rubric

## Student Version of the Rubric

### 5.3.2 Authority

Respond to the following prompts in the space below, using complete sentences:

- Identify the "domain type" of the site you're evaluating and explain why that is acceptable or unacceptable for your needs.
- Identify the "publisher" or host of the site and tell what you know (or can find out) about it.
- State whether or not the site is a personal site and explain why that is acceptable or unacceptable for your needs.
- State who (name the person or institution) created the site and tell what you know (or can find out) about the creator.
- Look for the author's credentials on the site. List his/her credentials and draw conclusions based on those credentials. If there are no credentials listed, tell what conclusions you can draw from their absence.
- Using what you know about the AUTHORITY of this web site, explain why it is or is not appropriate to use for your paper/project.

Evaluation Criteria	Beginning	Developing	Exemplary
<b>Uses Criteria Terms</b>	0 - This is the score you would receive if you don't address the authority of the web site you're evaluating.	1 - This is the score you would receive if you address the authority of the web site you're evaluating, but you don't actually use precise terminology in your answer, such as: "authority", "sponsorship", or "authorship".	2 - This is the score you would receive if you address the authority of the web site you're evaluating and you use precise terminology such as: "authority", "sponsorship", or "authorship".
<b>Cites Clues/Indicators of Criteria</b>	0 - This is the score you would receive if you don't address the signs or "indicators" of authority you looked for in the web site you're evaluating.	1 - This is the score you would receive if you address the signs or "indicators" of authority that you looked for in the web site you're evaluating, but you don't name specific indicators.	2 - This is the score you would receive if you address the signs or "indicators" of authority that you looked for in the web site and named them, such as: URL, "tip offs" (domain, server, ~), presence of personal or corporate author name, email, "About Us" or "Contact Us" links, or credentials.
<b>Cites Examples from Source</b>	0 - This is the score you would receive if you don't provide examples of authority indicators you looked for from the web site you're evaluating.	1 - This is the score you would receive if you refer vaguely to examples of authority indicators, but you don't cite specific examples from the web site you're evaluating.	2 - This is the score you would receive if you cite specific examples of authority indicators from the web site you're evaluating.
<b>Judges Whether or Not To Use Source</b>	0 - This is the score you would receive if you don't state whether or not the web site you're evaluating is appropriate to use for your assignment.	1 - This is the score you would receive if you state whether or not the web site you're evaluating is appropriate to use for your assignment, but you don't explain your reasoning for that decision.	2 - This is the score you would receive if you indicate whether or not the web site you're evaluating is appropriate to use for your assignment and explain your reasoning for that decision.

## Full Rubric

### 5.3.2 Authority

Evaluation Criteria	Beginning	Developing	Exemplary	Student Learning Outcomes
<b>Articulates Criteria</b>	0 - Student does not address authority issues.	1 - Student addresses authority issues, but does not use criteria terminology such as: author, authority, authorship, or sponsorship.	2 - Student addresses authority issues and uses criteria terminology such as: author, authority, authorship, or sponsorship.	<b>LOBO 3.1.1</b> The student will articulate established evaluation criteria. (ACRL 3.2.a)
<b>Cites Indicators of Criteria</b>	0 - Student does not address authority indicators.	1 - Student refers vaguely or broadly to authority indicators, but does not cite specific indicators such as: domain, server, or ~ in URL; presence of personal or corporate author name, email, "About Us" or "Contact Us" links; or author credentials.	2 - Student cites specific authority indicators such as: domain, server, or ~ in URL; presence of personal or corporate author name, email, "About Us" or "Contact Us" links; or author credentials.	<b>LOBO 3.1</b> The student will apply criteria to analyze information, including: authority...to information and its source (ACRL 3.2.a, 3.2.c)
<b>Links Indicators to Examples from Source</b>	0 - Student does not cite 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.	<b>LOBO 3.1</b> The student will apply criteria to analyze information, including: authority...to information and its source (ACRL 3.2.a, 3.2.c) <b>LOBO 3.1.2</b> The student will investigate an author's qualifications and reputation. (ACRL 3.2.a)
<b>Judges Whether or Not To Use Source</b>	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.	<b>LOBO 3.2</b> The student will evaluate sources (e.g., article, web site, book, journal, database, catalog) for use. (ACRL 3.4.g) <b>LOBO 3.2.2</b> The student will indicate whether or not a specific, individual source (e.g., a particular web site, article, book) is appropriate for the purpose at hand and provide a rationale for that decision based on established evaluation criteria (see LOBO 3.1). (ACRL 3.2.c)

## What we want to learn...

- Can different groups of raters use our rubric to score student work consistently in the area of information literacy?
- Do raters give scores that are the same as scores assigned by the researcher?
- To what degree are students able to use authority as a criterion to evaluate a website?

### Keep in mind...

- You can only score what you can see.
- The rubric does not score writing skills.
- The rubric does not score the web site they chose to evaluate.



## **APPENDIX E**

### **Selected External Rater Materials**

#### **Inventory of Enclosed Materials**

##### **Read First**

- ❑ Context for Study (purple) – Explanation of the study and directions for participating.
- ❑ Inventory of Enclosed Materials (salmon) – This page.

##### **Materials to Return to Megan**

- ❑ UNC Consent Form (blue) – There are two UNC consent forms included in your materials. Sign and return one. You may wish to keep the other for your records.
- ❑ NCSU Consent Form (pink) – There are two NCSU consent forms included in your materials. Sign and return one. You may wish to keep the other for your records.
- ❑ Data for Describing Rater Groups (green) – Information you provide on this sheet may be used to describe the rater pool.
- ❑ “Practice Test” Student Responses (ivory) & Directions (ivory) – This practice test needs to be completed before you score the 75 student responses. It is important that your score sheets are complete. You need to rank each student as “beginning”, “developing”, or “exemplary” for all 4 criteria on the rubric. Missing scores will cause problems with data analysis.
- ❑ 75 Student Responses (white) & Directions (white) – This is the “meat” of this study. The 75 student responses should be scored in one sitting (if possible) after you complete the practice test and before you complete the rater survey. It is important that your score sheets are complete. You need to rank each student as “beginning”, “developing”, or “exemplary” for all 4 criteria on the rubric. Missing scores will cause problems with data analysis.
- ❑ Rater Survey (yellow) – This survey seeks to collect your thoughts on the rubric and scoring process. Please feel free to include extensive comments if you wish!
- ❑ Return Mail Checklist (goldenrod) – Please complete this checklist before returning materials to Megan.
- ❑ Postage Paid Return Mail Envelope

## Supporting Materials

Review these supporting materials if you wish to do so. We are attempting to model the ways in which a rubric might be used when it is exported to other campuses. In this scenario, it is likely that you would have access to some printed and online materials, and we have provided those for you to examine. It is up to you how deeply you delve into these items.

- ❑ LOBO splash page and URL – For reference. Use this URL if you'd like a closer look at the LOBO tutorial. Login as "Guest".
- ❑ Photocopies of LOBO tutorial module on web site evaluation – For reference. This section of LOBO includes the prompt that generated the student responses scored in this study.
- ❑ Scoring Rubric – Full Version – This version of the rubric includes the student learning outcomes on which each criteria of the scoring rubric is based. Due to space limitations, these outcomes are omitted from the rubric on the score sheets for this study.
- ❑ Scoring Rubric – Student Version – This version of the rubric is available to students as they work through the LOBO tutorial. It is the same as the rubric listed above in content, but the language is more accessible. Web logs show that no or nearly no students accessed the link to this rubric.
- ❑ LOBO Objectives & Outcomes – For reference. This document lists the outcomes that LOBO addresses and on which the scoring rubric is based. It is based on the *ACRL Information Literacy Competency Standards for Higher Education* and the *Objectives for Information Literacy Instruction: A Model Statement for Academic Librarians*.
- ❑ *ACRL Information Literacy Competency Standards for Higher Education* – For reference.

# READ ME FIRST!!

## Context for Study & Directions

### **Study Title:**

The Assessment of Student Information Literacy Skills: A Rubric Approach

### **Purpose of the Study:**

This study is predicated on the need of academic librarians to find a new approach to the assessment of information literacy skills. Prior approaches including satisfaction surveys and input/output measures may not adequately provide librarians with information about what students know and can do. Other current approaches including standardized multiple-choice tests and large-scale performance assessments also may fail to provide the data librarians need to improve instruction on a local level. Faced with calls for accountability and the sincere desire to improve student learning, librarians require a new approach to library instruction assessment.

At North Carolina State University, librarians are interested in the variety of tools available for use in the assessment of information literacy skills. Recently, we have focused on rubrics as an option for such assessments. We have created a limited number of rubrics based on the standards approved by the Association of College and Research Libraries (ACRL) and want to investigate how well these rubrics can be used to assess specific approaches to information literacy instruction.

In this study, we intend to investigate the reliability of a rubric used to score student responses to one of numerous open-ended questions embedded in our Library Online Basic Orientation (LOBO) tutorial (see enclosed LOBO screen shots). The LOBO tutorial is based on ACRL standards (see enclosed LOBO Objectives and Outcomes document) and is a mandatory requirement in English 101, a course taken by 97% of incoming NCSU first-year students.

The section of LOBO selected as a focus for this study covers web site evaluation. Students are asked to answer several questions about a web site they are considering as a source for an English 101 assignment. In this study, we take a closer look at the student responses to questions focused on the authority of the web site they select.

It is important to note that while the rubric used in this study seeks to describe the behavior one might expect from a student who is learning to assess the authority of a web site, it is not customized to the specific questions students are asked to answer in the LOBO tutorial. Because we want the rubric to be useful for assessing a variety of artifacts of student learning (e.g., research journals, worksheets, think aloud tasks), we eschewed the task-specific elements of LOBO and based our rubric on the *ACRL Information Literacy Competency Standards*

*for Higher Education and the Objectives for Information Literacy Instruction: A Model Statement for Academic Librarians.* We hope that this approach has allowed us to craft a rubric that can be used to assess student skills manifested in different ways, not one that would be limited to assessment of LOBO alone.

### **Study Questions:**

1. To what degree can different groups of raters provide consistent scoring of artifacts of student learning using a rubric?
2. To what degree can raters provide scores that are consistent with scores assigned by the researcher?

If the rubric is determined to be reliable and valid, then the study will also examine:

3. To what degree are students able to use authority as a criterion to evaluate a website?

The data resulting from this survey will be used to draw conclusions and make recommendations regarding the usefulness of rubrics as a tool for information literacy instruction assessment.

### **The Study at NCSU:**

Fifteen raters at North Carolina State University have already completed this study. Five librarians, five English 101 instructors, and five former English 101 students have completed the study following a 3 hour training and “norming” session where they were familiarized with the concept of rubric assessment and practiced coming to consensus on how to score student responses using the study rubric.

### **Where You Fit In:**

This study seeks to investigate not only the use of rubrics to assess information literacy on individual campuses where detailed explanation and training is possible, but also how an information literacy rubric might fare when it is exported to other campuses and libraries. Your participation will help us learn to what degree a rubric designed to measure information literacy skills can be used reliably by instruction librarians and reference librarians on campuses beyond our own who have not had the chance to talk about the rubric with others or be trained in its use. In a profession like ours where the sharing of strategies and techniques across campuses is so important, the data you will provide is of great value!

**Directions:**

1. Use the Inventory of Enclosed Materials (salmon) to ensure that you have all necessary study materials.
2. Read and sign one copy of the NCSU Consent Form (pink) and the UNC Consent Form (blue)
3. Complete the Data for Describing Rater Groups (green) sheet.
4. Review supporting materials if you wish to do so. We are attempting to model the ways in which a rubric might be used when it is exported to other campuses. In this scenario, it is likely that you would have access to some printed and online materials, and we have provided those for you to examine. It is up to you how deeply you delve into these items.
5. Read the directions for the “Practice Test” Student Responses. Then, score each of the “Practice Test” Student Responses (ivory). It is important that your score sheets are complete. You need to rank each student as “beginning”, “developing”, or “exemplary” for all 4 criteria on the rubric. Missing scores will cause problems with data analysis.
6. Read the directions for the 75 Student Responses. Then, score each of the 75 Student Responses (white). It is important that your score sheets are complete. You need to rank each student as “beginning”, “developing”, or “exemplary” for all 4 criteria on the rubric. Missing scores will cause problems with data analysis.
7. Complete the Rater Comment Sheet (yellow).
8. Complete the Return Mail Checklist (goldenrod) and return materials to Megan.

### **Data for Describing Rater Groups**

By completing the form below, you will provide the researcher with information that can be used to describe the rater groups in this study. Your name will not be used in conjunction with this data. In addition, this data will be kept in a secure location and viewed only by the researcher.

Rater # \_\_\_\_\_

#### **Non-NCSU Librarians**

Gender:

Race/Ethnicity:

Job Title:

Institution:

Number of years as a librarian:

Did/do you contribute to the creation or maintenance of an information literacy tutorial? If so, describe briefly.

Are you responsible for library instruction? If so, describe briefly.

Are you responsible for assessment of library instruction? If so, describe briefly.

Did/do you use rubrics for assessment purposes? If so, describe briefly.

## Directions for Scoring the “Practice Test”

Attached are 10 student responses to the open-ended questions asked in LOBO and the rubrics for scoring them. **These 10 score sheets make up the “Practice Test” that will be used to learn about the scores a rater provides when s/he first begins scoring.**

On each sheet, you'll find a student response, the rubric you use to score the response, and at the bottom are the questions asked in LOBO for easy reference. **Please be sure to score each student completely. Place a √ or X in the circle that corresponds to the appropriate performance level (“Beginning”, “Developing”, or “Exemplary”) for each of the 4 evaluation criteria listed along the left side of the rubric.** If you wish, you may provide a total score at the bottom right of the rubric, as in the example below. (Note: This total score is *not* intended to serve as a “grade”.)

Evaluation Criteria	Beginning	Developing	Exemplary
<b>Articulates Criteria</b>	0 - Student does not address authority issues. <div style="text-align: right;">○</div>	1 - Student addresses authority issues, but does not use criteria terminology. <div style="text-align: right;">○</div>	2 - Student addresses authority issues and uses criteria terminology such as: author, authority, authorship, or sponsorship. <div style="text-align: right;">○</div>
<b>Cites Indicators of Criteria</b>	0 - Student does not address authority indicators. <div style="text-align: right;">○</div>	1 - Student refers vaguely or broadly to authority indicators, but does not cite specific indicators. <div style="text-align: right;">○</div>	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. <div style="text-align: right;">○</div>
<b>Links Indicators to Examples from Source</b>	0 - Student does not address examples of authority indicators from the site. <div style="text-align: right;">○</div>	1 - Student refers vaguely or broadly to examples of authority indicators from the site under consideration, but does not cite specific examples. <div style="text-align: right;">○</div>	2 - Student cites specific examples of authority indicators from the site under consideration. <div style="text-align: right;">○</div>
<b>Judges Whether or Not To Use Source</b>	0 - Student does not indicate whether or not the site is appropriate to use for the purpose at hand. <div style="text-align: right;">○</div>	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. <div style="text-align: right;">○</div>	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. <div style="text-align: right;">○</div>

RESEARCHER USE ONLY: Total Score \_\_\_\_/8

Keep the following in mind as you score student responses.

- You can only score what you can see.
- The rubric does not score writing skills.
- The rubric does not score the website the student chose to evaluate.

### **Before returning these scores...**

Please double-check to be sure that your score sheets are complete. You need to rank each student as “beginning”, “developing”, or “exemplary” for all 4 criteria.

**Missing scores will cause problems with data analysis.**

## Directions for Scoring 75 Student Responses

Attached are the 75 student responses to the open-ended questions asked in LOBO and the rubrics for scoring them. On each sheet, you'll find a student response, the rubric you use to score the response, and at the bottom are the questions asked in LOBO for easy reference. **Please be sure to score each student completely. Place a  $\checkmark$  or X in the circle that corresponds to the appropriate performance level ("Beginning", "Developing", or "Exemplary") for each of the 4 evaluation criteria listed along the left side of the rubric.** If you wish, you may provide a total score at the bottom right of the rubric, as in the example below. (Note: This total score is *not* intended to serve as a "grade".)

Evaluation Criteria	Beginning	Developing	Exemplary
<b>Articulates Criteria</b>	0 - Student does not address authority issues. <div style="text-align: right;"><input type="radio"/></div>	1 - Student addresses authority issues, but does not use criteria terminology. <div style="text-align: right;"><input type="radio"/></div>	2 - Student addresses authority issues and uses criteria terminology such as: author, authority, authorship, or sponsorship. <div style="text-align: right;"><input type="radio"/></div>
<b>Cites Indicators of Criteria</b>	0 - Student does not address authority indicators. <div style="text-align: right;"><input type="radio"/></div>	1 - Student refers vaguely or broadly to authority indicators, but does not cite specific indicators. <div style="text-align: right;"><input type="radio"/></div>	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. <div style="text-align: right;"><input type="radio"/></div>
<b>Links Indicators to Examples from Source</b>	0 - Student does not address examples of authority indicators from the site. <div style="text-align: right;"><input type="radio"/></div>	1 - Student refers vaguely or broadly to examples of authority indicators from the site under consideration, but does not cite specific examples. <div style="text-align: right;"><input type="radio"/></div>	2 - Student cites specific examples of authority indicators from the site under consideration. <div style="text-align: right;"><input type="radio"/></div>
<b>Judges Whether or Not To Use Source</b>	0 - Student does not indicate whether or not the site is appropriate to use for the purpose at hand. <div style="text-align: right;"><input type="radio"/></div>	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. <div style="text-align: right;"><input type="radio"/></div>	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. <div style="text-align: right;"><input type="radio"/></div>

RESEARCHER USE ONLY: Total Score \_\_\_\_/8

Keep the following in mind as you score student responses.

- You can only score what you can see.
- The rubric does not score writing skills.
- The rubric does not score the website the student chose to evaluate.

Please score all rubrics in one sitting and enter your start and ending times below.

Start Time: \_\_\_\_\_ End Time: \_\_\_\_\_

### Before returning these scores...

Please double-check to be sure that your score sheets are complete. You need to rank each student as "beginning", "developing", or "exemplary" for all 4 criteria.

**Missing scores will cause problems with data analysis.**



## APPENDIX F

### Student Version of Study Rubric

Evaluation Criteria	Beginning	Developing	Exemplary
<b>Uses Criteria Terms</b>	0 - This is the score you would receive if you don't address the authority of the web site you're evaluating.	1 - This is the score you would receive if you address the authority of the web site you're evaluating, but you don't actually use precise terminology in your answer, such as: "authority", "sponsorship", or "authorship".	2 - This is the score you would receive if you address the authority of the web site you're evaluating and you use precise terminology such as: "authority", "sponsorship", or "authorship".
<b>Cites Clues/Indicators of Criteria</b>	0 - This is the score you would receive if you don't address the signs or "indicators" of authority you looked for in the web site you're evaluating.	1 - This is the score you would receive if you address the signs or "indicators" of authority that you looked for in the web site you're evaluating, but you don't name specific indicators.	2 - This is the score you would receive if you address the signs or "indicators" of authority that you looked for in the web site and named them, such as: URL "tip offs" (domain, server, ~), presence of personal or corporate author name, email, "About Us" or "Contact Us" links, or credentials.
<b>Cites Examples from Source</b>	0 - This is the score you would receive if you don't provide examples of authority indicators you looked for from the web site you're evaluating.	1 - This is the score you would receive if you refer vaguely to examples of authority indicators, but you don't cite specific examples from the web site you're evaluating.	2 - This is the score you would receive if you cite specific examples of authority indicators from the web site you're evaluating.
<b>Judges Whether or Not To Use Source</b>	0 - This is the score you would receive if you don't state whether or not the web site you're evaluating is appropriate to use for your assignment.	1 - This is the score you would receive if you state whether or not the web site you're evaluating is appropriate to use for your assignment, but you don't explain your reasoning for that decision.	2 - This is the score you would receive if you indicate whether or not the web site you're evaluating is appropriate to use for your assignment and explain your reasoning for that decision.

## APPENDIX G

### UNC Consent Form

University of North Carolina-Chapel Hill  
Consent to Participate in a Research Study  
Adult Participants  
Social Behavioral Form

---

IRB Study # 05-053  
Consent Form Version Date: 9-6-05

**Title of Study:** The Assessment of Student Information Literacy Skills: A Rubric Approach

**Principal Investigator:** Megan Oakleaf  
**UNC-Chapel Hill Department:** School of Information and Library Science  
**UNC-Chapel Hill Phone number:** 919-513-0346  
**Email Address:** megan\_oakleaf@ncsu.edu  
**Faculty Advisor:** Dr. Helen Tibbo  
**Funding Source:** NCSU CUPR Mini-Grant

**Study Contact telephone number:** 919-513-0346  
**Study Contact email:** megan\_oakleaf@ncsu.edu

---

#### **What are some general things you should know about research studies?**

You are being asked to take part in a research study. To join the study is voluntary.

You may refuse to join, or you may withdraw your consent to be in the study, for any reason. If you refuse to join or withdraw from the study, you will not receive payment for participation.

Research studies are designed to obtain new knowledge. This new information may help people in the future. As a study participant, you may benefit from learning new information, but learning is not guaranteed. It is possible that you may not receive any direct benefit from being in the research study.

Details about this study are discussed below. It is important that you understand this information so that you can make an informed choice about being in this research study.

You will be given a copy of this consent form. You should ask the researchers named above, or staff members who may assist them, any questions you have about this study at any time.

**What is the purpose of this study?**

The purpose of this research study is to learn about ways to measure students' information literacy skills. This study seeks to examine a different approach to assessing student skills—a rubric approach.

**How many people will take part in this study?**

If you decide to be in this study, you will be one of approximately 25 people in this research study.

**How long will your part in this study last?**

Your part of this study is estimated to last 6-8 hours.

**What will happen if you take part in the study?**

If you participate in this study, you will learn about a rubric designed to assess students' ability to evaluate a website based on its authority.

You will receive information regarding the rubric and directions about how to use the rubric.

You will then score 75 student responses using the rubric. Scoring these responses is a requirement of participating in this study. No other raters will see the scores you assign to these responses.

*You will finish by completing a short true or false survey about your experience using the rubric. You may choose not to answer a question on the survey for any reason.*

**What are the possible benefits from being in this study?**

Research is designed to benefit society by gaining new knowledge. You may not benefit personally from being in this research study.

**What are the possible risks or discomforts involved from being in this study?**

There are no known or anticipated risks to being in this study. There may be uncommon or previously unknown risks. You should report any problems to the researcher.

**How will your privacy be protected?**

Your privacy will be protected by assigning you a number so that your name does not appear on survey materials. The key to names and numbers will be kept on a password-protected server and a locked file cabinet. The survey materials will also be kept in a locked cabinet. Only the primary investigator will have access to individually identifiable data.

Participants will not be identified in any report or publication about this study. Although every effort will be made to keep research records private, there may

be times when federal or state law requires the disclosure of such records, including personal information. This is very unlikely, but if disclosure is ever required, UNC-Chapel Hill will take steps allowable by law to protect the privacy of personal information. In some cases, your information in this research study could be reviewed by representatives of the University, research sponsors, or government agencies for purposes such as quality control or safety.

**Will you receive anything for being in this study?**

Participants who attend on-campus training for this study will receive \$50.00. Participants who score student responses will receive \$75.00. Payments will be distributed at the end of your participation in the study.

**Will it cost you anything to be in this study?**

There will be no costs for being in the study.

**What if you have questions about this study?**

You have the right to ask, and have answered, any questions you may have about this research. If you have questions, or concerns, you should contact the researchers listed on the first page of this form.

**What if you have questions about your rights as a research participant?**

All research on human volunteers is reviewed by a committee that works to protect your rights and welfare. If you have questions or concerns about your rights as a research subject you may contact, anonymously if you wish, the Institutional Review Board at 919-966-3113 or by email to IRB\_subjects@unc.edu.

-----  
**Participant's Agreement:**

I have read the information provided above. I have asked all the questions I have at this time. I voluntarily agree to participate in this research study.

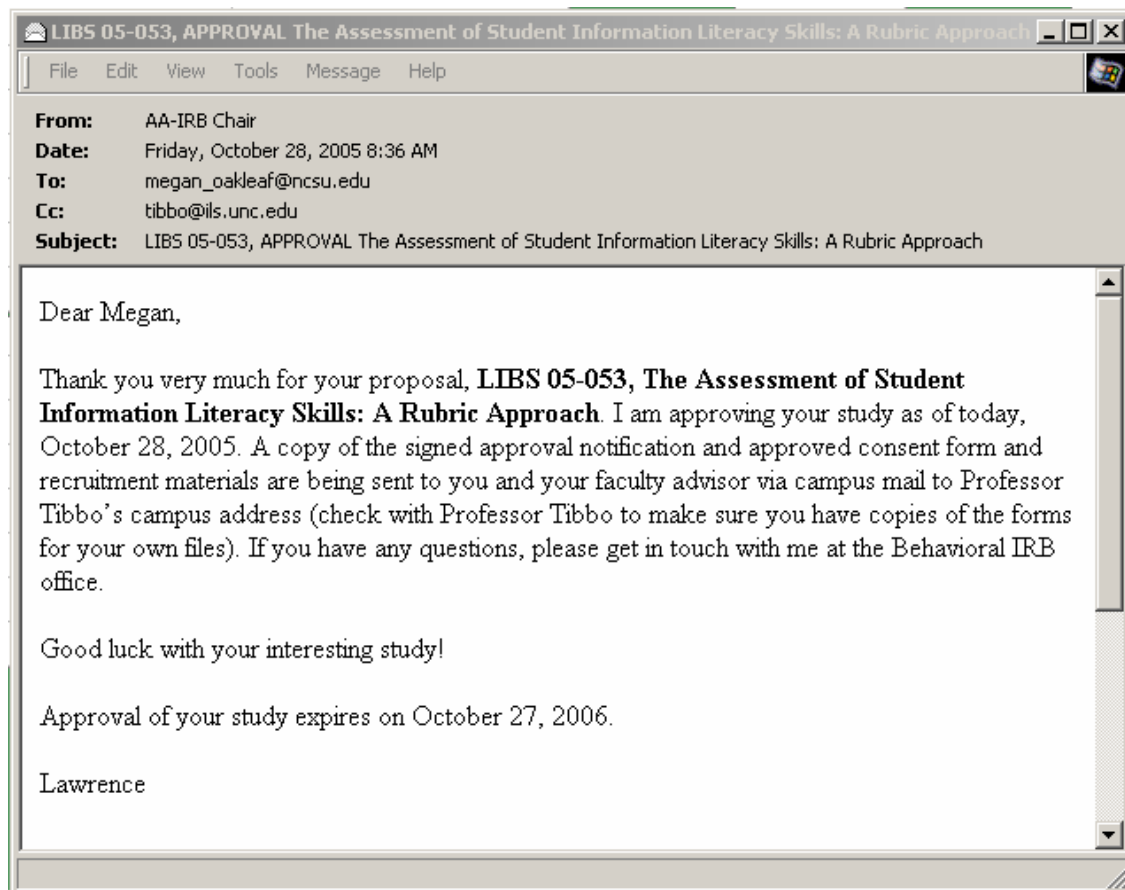
\_\_\_\_\_  
Signature of Research Participant

\_\_\_\_\_  
Date

\_\_\_\_\_  
Printed Name of Research Participant

## APPENDIX H

### UNC IRB Approval



## **APPENDIX I**

### **NCSU Consent Form**

#### **North Carolina State University INFORMED CONSENT FORM for RESEARCH**

##### **The Assessment of Student Information Literacy Skills: A Rubric Approach**

Megan Oakleaf

---

We are asking you to participate in a research study. The purpose of this research study is to learn about ways to measure students' information literacy skills. There are few viable options for assessing students' ability to find, evaluate, and use information. Those that exist have drawbacks. This study seeks to examine a different approach to assessing student skills—a rubric approach. Rubric assessment is used successfully in other areas, most notably writing assessment. If studies can demonstrate that rubrics can be used consistently to measure information literacy skills, they might offer a new option to librarians and instructors seeking to teach and build these skills in their students.

#### **INFORMATION**

If you participate in this study, you will learn about a rubric designed to assess students' ability to evaluate a website based on its authority.

You will receive information regarding the rubric and directions about how to use the rubric.

You will then score 75 student responses using the rubric. Scoring these responses is a requirement of participating in this study. No other raters will see the scores you assign to these responses.

You will finish by completing a short true or false survey about your experience using the rubric. You may choose not to answer a question on the survey for any reason.

Your part of this study is estimated to last 6-8 hours.

#### **RISKS**

**There are no known or anticipated risks to being in this study.**

#### **BENEFITS**

Research is designed to benefit society by gaining new knowledge. You may not benefit personally from being in this research study.

#### **CONFIDENTIALITY**

The information in the study records will be kept strictly confidential. Your privacy will be protected by assigning you a number so that your name does not appear on survey materials. The key to names and numbers will be kept on a password-protected server and a locked file cabinet. The survey materials will also be kept in a locked cabinet. Only the primary investigator will have access to individually identifiable data. No reference will be made in oral or written reports which could link you to the study.

#### **COMPENSATION (if applicable)**

For participating in this study you will receive \$50.00 for onsite training involved with this study (if you are in the group that receives training), and you will be receiving \$75.00 for scoring student responses as a part of this study. Both payments will be distributed after the end of your participation in the study. If you do not complete the scoring of student responses, you will not be paid for that portion of the study.

#### **CONTACT**

**If you have questions at any time about the study or the procedures, you may contact the researcher, Megan Oakleaf, at 2205 Hillsborough St. NCSU Libraries Raleigh, NC 27695, or (919) 513-0346. If you feel you have not been treated according to the descriptions in this form, or your rights as a participant in research have been violated during the course of this project, you may contact Dr. Matthew Zingraff, Chair of the NCSU IRB for the Use of Human Subjects in Research Committee, Box 7514, NCSU Campus (919/513-1834) or Mr. Matthew Ronning, Assistant Vice Chancellor, Research Administration, Box 7514, NCSU Campus (919/513-2148)**

**PARTICIPATION**

Your participation in this study is voluntary; you may decline to participate without penalty. If you decide to participate, you may withdraw from the study at any time without penalty and without loss of benefits to which you are otherwise entitled. If you withdraw from the study before data collection is completed your data will be returned to you or destroyed at your request.

**CONSENT**

**“I have read and understand the above information. I have received a copy of this form. I agree to participate in this study with the understanding that I may withdraw at any time.”**

**Subject's signature**\_\_\_\_\_ **Date** \_\_\_\_\_

**Investigator's signature**\_\_\_\_\_ **Date** \_\_\_\_\_

## **APPENDIX J**

### **NCSU IRB Approval**

Sponsored Programs and  
Regulatory Compliance  
Campus Box 7514  
1 Leazar Hall  
Raleigh, NC 27695-7514  
919.515.7200  
919.515.7721 (fax)

**From:** Debra A. Paxton, Regulatory Compliance Administrator  
North Carolina State University  
Institutional Review Board

**Date:** September 19, 2005

**Project Title:** The assessment of student information literacy skills: A rubric approach

**IRB#:** 193-05-9

Dear Dr. Oakleaf:

The research proposal named above has received administrative review and has been approved as exempt from the policy as outlined in the Code of Federal Regulations (Exemption: 46.101.b.6). Provided that the only participation of the subjects is as described in the proposal narrative, this project is exempt from further review.

**NOTE:**

1. This committee complies with requirements found in Title 45 part 46 of The Code of Federal Regulations. For NCSU projects, the Assurance Number is: FWA00003429; the IRB Number is: IRB00000330
2. Review de novo of this proposal is necessary if any significant alterations/additions are made.

Thank you.  
Sincerely,

Debra Paxton  
NCSU IRB



## WORKS CITED

- Alliance for Childhood. Tech Tonic: Towards a New Literacy of Technology. College Park, Maryland: Alliance for Childhood, 2004.
- American Association of Higher Education. "Nine Principles of Good Practice for Assessing Student Learning". 1996. 28 Feb. 2005.  
<<http://www.aahe.org/assessment/principi.htm>>.
- American Library Association. "Information Literacy Competency Standards for Higher Education". 2000. April 22 2005.  
<<http://www.ala.org/ala/acrl/acrlstandards/informationliteracycompetency.htm>>.
- . "Presidential Committee on Information Literacy: Final Report". 1989. May 15 2005. <<http://www.ala.org/ala/acrl/acrlpubs/whitepapers/presidential.htm>>.
- Andrade, Heidi Goodrich. "Using Rubrics to Promote Thinking and Learning." Educational Leadership 57.5 (2000): 13-18.
- Angelo, Thomas A. "Reassessing (and Defining) Assessment." AAHE Bulletin 48.3 (1995): 7.
- Arter, Judith A. "Using Assessment as a Tool for Learning." A Handbook for Student Performance in an Era of Restructuring. Eds. Robert E. Blum and Judith A. Arter. Alexandria, Virginia: Association for Supervision and Curriculum Development, 1996. IV-10:1 - IV-10:6.
- Arter, Judith, and Jay McTighe. Scoring Rubrics in the Classroom: Using Performance Criteria for Assessing and Improving Student Performance. Thousand Oaks, California: Corwin Press, 2000.
- Asaravala, Amit. "Testing Your Tech Smarts." Wired News April 8 2005.
- "Assessment Methods". 1999. Assessment Strategies Inc.  
<<http://www.asinc.ca/en/methods/index.html>>.
- Association of College and Research Libraries. "Introduction to Information Literacy". 2003. May 15 2005.  
<<http://www.ala.org/ala/acrl/acrlissues/acrlinfolit/infolitoverview/introtoinfolit/introinfolit.htm>>.
- . "Objectives for Information Literacy Instruction: A Model Statement for Academic Librarians". 2001. 25 March 2005.  
<<http://www.ala.org/ala/acrl/acrlstandards/objectivesinformation.htm>>.

- Bagnato, Kristin. "New Exam Tests Students' IT Abilities." Community College Week (2005).
- Baker, Eva L., et al. "Dimensionality and Generalizability of Domain-Independent Performance Assessments." Journal of Education Research 89.4 (1995): 197-205.
- Baker, Eva L., and Harold F. O'Neil Jr. "Standards for Student Learning." Encyclopedia of Education. Ed. James W. Guthrie. Vol. 6. New York: Macmillan Reference USA, 2003. 2315-18.
- Baker, Neal, and Steve McKinzie. "Librarians and Faculty in Tandem: Taking Our Cue from the Evening News." Reference and User Services Quarterly 37.1 (1997): 19.
- Barclay, Donald. "Evaluating Library Instruction: Doing the Best You Can with What You Have." Research Quarterly 33.2 (1993): 195-203.
- Battersby, Mark. "So, What's a Learning Outcome Anyway?" Vancouver, B.C: Centre for Curriculum Transfer and Technology, 1999.
- . "So, What's a Learning Outcome Anyway? Learning Outcomes and the Learning Paradigm". 2002. 2004.  
<<http://merlin.capcollege.bc.ca/mbatters/whatsalearningoutcome.htm>>.
- Beck, Susan E., and Nancy B. Turner. "On the Fly BI: Reaching and Teaching from the Reference Desk." Reference Librarian 72 (2001): 83-96.
- Bernier, Rosemarie. "Making Yourself Indispensable by Helping Teachers Create Rubrics." CSLA Journal 27.2 (2004): 24-25.
- Blandy, Susan Griswold. "The Librarians' Role in Academic Assessment and Accreditation: A Case Study." Assessment and Accountability in Reference Work New York Haworth Press, 1992. 69-87.
- Blessing, Laura. "Gender and Race in the NCSU Libraries." E-mail to Megan Oakleaf. 10 August 2005.
- Blixrud, Julia C. "Establishing a Role for Research Libraries in Learning Outcomes Assessment". 2000. A Bimonthly Report on Research Library Issues and Actions from ARL, CNI, and SPARC. (December 2000): Association of Research Libraries.  
<<http://www.arl.org/newsltr/213/assess.html>>.
- Bresciani, Marilee J. "Outcomes Assessment in Student Affairs: Moving Beyond Satisfaction to Student Learning and Development". 2002. Net Results.

- National Association of Student Personnel Administrators. 19 Feb. 2005.  
<<http://www.naspa.org/membership/mem/nr/article.cfm?id=884>>.
- Bresciani, Marilee J., Carrie L. Zelna, and James A. Anderson. Assessing Student Learning and Development: A Handbook for Practitioners. National Association of Student Personnel Administrators, 2004.
- Brookhart, Susan M. "A Theoretical Framework for the Role of Classroom Assessment in Motivating Student Effort and Achievement." Applied Measurement in Education 10.2 (1997): 161-80.
- Buchanan, Lori E. "Assessing Liberal Arts Classes." Assessing Student Learning Outcomes for Information Literacy Instruction in Academic Institutions. Ed. Elizabeth Fuseler Avery. Chicago: Association of College and Research Libraries, 2003. 68-73.
- Callison, Daniel. "Rubrics." School Library Media Activities Monthly 17.2 (2000): 34-36, 42.
- Caravello, Patti S., et al. "Information Competence at UCLA: Report of a Survey Project". 2001. April 29 2005.  
<<http://www.library.ucla.edu/infocompetence/toc.html>>.
- Carter, Elizabeth W. "'Doing the Best You Can with What You Have': Lessons Learned from Outcomes Assessment." Journal of Academic Librarianship 28.1 (2002): 36-41.
- Choinski, Elizabeth, Amy E. Mark, and Missy Murphey. "Assessment with Rubrics: An Efficient and Objective Means of Assessing Student Outcomes in an Information Resources Class." Portal: Libraries and the Academy 3.4 (2003): 563-75.
- Colton, Dean A., et al. Reliability Issues with Performance Assessments: A Collection of Papers. Iowa City: ACT, 1997.
- Council for Higher Education Accreditation. "Student Learning Outcomes Workshop." CHEA Chronicle 5.2 (2002): 1-3.
- Council of Australian University Librarians. "CAUL Information Skills Survey". 2004 ANZIIL News. <<http://www.anziil.org/news/february2004.htm>>.
- Crighton, Johanna V. "Standardized Tests and Educational Policy." Encyclopedia of Education. Ed. James W. Guthrie. Vol. 7. New York: Macmillan Reference USA, 2003. 2530-33.

- D'Angelo, Barbara J. "Integrating and Assessing Information Competencies in a Gateway Course." Reference Services Review 29.4 (2001): 282-93.
- DeFranco, Francine, and Richard Bleiler. Spec Kit #279: Evaluating Library Instruction. Washington, D.C.: Association of Research Libraries, 2003.
- Dewald, Nancy H. "Transporting Good Library Instruction Practices into the Web Environment: An Analysis of Online Tutorials." Journal of Academic Librarianship 25.1 (1999): 26-31.
- Dugan, Robert E., and Peter Hernon. "Outcomes Assessment: Not Synonymous with Inputs and Outputs." Journal of Academic Librarianship 28.6 (2002): 376-80.
- Dunn, Kathleen. "Information Competence Assessment: Web-Based Assessment of University Entry-Level Information Competence". 1999. April 30 2005. <<http://www.csupomona.edu/~library/InfoComp/>>.
- Educational Testing Service. "ETS Collaborates with Major Universities to Assess 21st Century Skills". 2003. March 21, 2004. <<http://www.ets.org/news/03091701.html>>.
- . ICT Literacy Assessment: An Issue Paper from ETS, 2004.
- . "Succeeding in the 21st Century: What Higher Education Must Do to Address the Gap in Information and Communication Technology Proficiencies". 2003. 2004. <<http://www.calstate.edu/LS/ICTwhitepaperfinal.pdf>>.
- Elmborg, James K. "Teaching at the Desk: Toward a Reference Pedagogy." Portal: Libraries and the Academy 2.3 (2002): 455-64.
- Emmons, Mark, and Wanda Martin. "Engaging Conversation: Evaluating the Contribution of Library Instruction to the Quality of Student Research." College and Research Libraries 63.6 (2002): 545-60.
- Evans, Candy Duncan. "Understanding Assessment." Delta Pi Epsilon Journal 44.1 (2002): 3-12.
- Farmer, Lesley S. Johnson. "Authentic Assessment of Information Literacy through Electronic Products." Book Report 16 (1997): 11-13.
- Feinberg, Richard, and Christine King. "Performance Evaluation in Bibliographic Instruction Workshop Courses: Assessing What Students Do as a Measure of What They Know." Reference Services Review 20.2 (1992): 75-80.

- Flaspohler, Molly R. "Information Literacy Program Assessment: One Small College Takes the Big Plunge." Reference Services Review 31.2 (2003): 129-40.
- Foote, Chandra J., Paul J. Vermette, and Catherine F. Battaglia. Constructivist Strategies: Meeting Standards and Engaging Adolescent Minds. Larchmont, New York: Eye on Education, 2001.
- Fourie, Ina, and Daleen Van Niekerk. "Using Portfolio Assessment in a Module in Research Information Skills." Education for Information 17.4 (1999): 333-53.
- Franks, Dana. "Using Rubrics to Assess Information Literacy Attainment in a Community College Education Class." Assessing Student Learning Outcomes for Information Literacy Instruction in Academic Institutions. Ed. Elizabeth Fuseler Avery. Chicago: Association of College and Research Libraries, 2003. 132-47.
- Gabler, Ina Claire, and Michael Schroeder. Seven Constructivist Methods for the Secondary Classroom: A Planning Guide for Invisible Teaching. Boston: Allyn and Bacon, 2003.
- Gauss, Nancy, and Kathleen Kinkema. "Webliography Assignment for a Lifetime Wellness Class." Assessing Student Learning Outcomes for Information Literacy Instruction in Academic Institutions. Ed. Elizabeth Fuseler Avery. Chicago: Association of College and Research Libraries, 2003. 161-71.
- Graham, Sandra. "Motivation: Overview." Encyclopedia of Education. Ed. James W. Guthrie. Vol. 5. New York: Macmillian Reference USA, 2003. 1690-95.
- Grassian, Esther S., and Joan R. Kaplowitz. Information Literacy Instruction. New York: Neal-Schuman, 2001.
- Gratch Lindauer, Bonnie. "Comparing the Regional Accreditation Standards: Outcomes Assessment and Other Trends." Journal of Academic Librarianship 28.1 (2002): 14-25.
- . "Defining and Measuring the Library's Impact on Campuswide Outcomes." College and Research Libraries 59.6 (1998): 546-63.
- Gwet, Kilem. Handbook of Inter-Rater Reliability: How to Estimate the Level of Agreement between Two or Multiple Raters. Gaithersburg, Maryland: STATAXIS, 2001.

- Hafner, John C. "Quantitative Analysis of the Rubric as an Assessment Tool: An Empirical Study of Student Peer-Group Rating." International Journal of Science Education 25.12 (2003): 1509-28.
- Hall, Tracie. "Question About Minority Makeup of Library Profession." E-mail to Megan Oakleaf. 18 August 2005.
- Herman, John, Pamela R. Aschbacher, and Lynn Winters. "Setting Criteria." A Handbook for Student Performance in an Era of Restructuring. Eds. Robert E. Blum and Judith A. Arter. Alexandria, Virginia: Association for Supervision and Curriculum Development, 1996. VI-4:1 - VI-4:19.
- Hess, Stephen K. "Validating and Anchoring Classroom Performance Tasks." A Handbook for Student Performance in an Era of Restructuring. Eds. Robert E. Blum and Judith A. Arter. Alexandria, Virginia: Association for Supervision and Curriculum Development, 1996. V-9:1 - V-9:4.
- Hiller, Steve, and Jim Self. "Making Library Assessment Work." ARL 4th Human Resources Management Symposium. Washington, D.C., 2004.
- Horan, Mark. "What Students See: Sketch Maps as Tools for Assessing Knowledge of Libraries." Journal of Academic Librarianship 25.3 (1999): 187-201.
- Hovde, Karen. "Check the Citation: Library Instruction and Student Paper Bibliographies." Research Strategies 17 (2000): 3-9.
- Hoyler, Robert. "The Road Not Taken." Change 30.5 (1998).
- Huba, Mary E., and Jann E. Freed. Learner-Centered Assessment on College Campuses. Boston: Allyn and Bacon, 2000.
- Hutchins, Elizabeth O. "Assessing Student Learning Outcomes in Political Science Classes." Assessing Student Learning Outcomes for Information Literacy Instruction in Academic Institutions. Ed. Elizabeth Fuseler Avery. Chicago: Association of College and Research Libraries, 2003. 172-84.
- Iannuzzi, Patricia. "We Are Teaching, but Are They Learning: Accountability, Productivity, and Assessment." Journal of Academic Librarianship 25.4 (1999): 304-05.
- International ICT Literacy Panel. "Digital Transformation: A Framework for ICT Literacy". 2002. Educational Testing Service 2004.  
<<http://www.ets.org/research/ictliteracy/>>.

- Johnson, Robert L., James Penny, and Belita Gordon. "The Relation between Score Resolution Methods and Interrater Reliability: An Empirical Study of an Analytic Scoring Rubric." Applied Measurement in Education 13.2 (2000): 121-38.
- Kasowitz-Scheer, Abby, and Michael Pasqualoni. "Information Literacy Instruction in Higher Education: Trends and Issues." ERIC Digest ED465375 (2002).
- Kilcullen, Maureen. "Teaching Librarians to Teach: Recommendations on What We Need to Know." Reference Services Review 26.2 (1998): 7-18.
- King, David N., and John C. Ory. "Effects of Library Instruction on Student Research: A Case Study." College and Research Libraries 42.1 (1981): 31-41.
- Kivel, Andy. "Institutionalizing a Graduation Requirement." Assessing Student Learning Outcomes for Information Literacy Instruction in Academic Institutions. Ed. Elizabeth Fuseler Avery. Chicago: Association of College and Research Libraries, 2003. 192-200.
- . "SF Bay Area Community College Information Competency Assessment Project." ACRL National Conference. Charlotte, North Carolina: Association of College and Research Libraries, 2003.
- Knight, Lorrie A. "Assessing Student Learning through the Analysis of Research Papers." Assessing Student Learning Outcomes for Information Literacy Instruction in Academic Institutions. Ed. Elizabeth Fuseler Avery. Chicago: Association of College and Research Libraries, 2003. 201-06.
- . "The Role of Assessment in Library User Education." Reference Services Review 30.1 (2002): 15-24.
- Kobritz, Barbara. "Information Literacy in Community College Communications Courses." Assessing Student Learning Outcomes for Information Literacy Instruction in Academic Institutions. Ed. Elizabeth Fuseler Avery. Chicago: Association of College and Research, 2003. 207-15.
- Kuh, George D., and Robert M. Gonyea. "The Role of the Academic Library in Promoting Student Engagement in Learning." College and Research Libraries 64.4 (2003): 256-82.
- Lamon, Mary. "Constructivist Approach." Encyclopedia of Education. Ed. James W. Guthrie. Vol. 4. New York: Macmillan Reference USA, 2003. 1463-67.

- Landis, J. Richard, and Gary G. Koch. "The Measure of Observer Agreement for Categorical Data." Biometrics 33 (1977): 159-74.
- Lawson, Mollie D. "Assessment of a College Freshman Course in Information Resources." Library Review 48.2 (1999): 73-79.
- Lazzaro, William L. "Empowering Students with Instructional Rubrics." A Handbook for Student Performance in an Era of Restructuring. Eds. Robert E. Blum and Judith A. Arter. Alexandria, Virginia: Association for Supervision and Curriculum Development, 1996. VI-3:1 - VI-3:9.
- Lepper, Mark R., and Jennifer Henderlong. "Motivation: Instruction." Encyclopedia of Education. Ed. James W. Guthrie. Vol. 5. New York: Macmillan Reference USA, 2003. 1695-98.
- Lichtenstein, Art A. "Informed Instruction: Learning Theory and Information Literacy." Journal of Educational Media and Library Sciences 38.1 (2000): 22-31.
- Lincoln, Yvonna. "Authentic Assessment and Research Methodology." E-mail to Megan Oakleaf. 2005.
- MacColl, Gail S., and Kathleen D. White. "Communicating Educational Research Data to General, Nonresearcher Audiences." Practical Assessment, Research, and Evaluation 6.7 (1998).
- Maki, Peggy L. Assessing for Learning: Building a Sustainable Commitment across the Institution. Sterling, Virginia: Stylus, 2004.
- . "From Standardized Tests to Alternative Methods: Assessing Learning in Education." Change 33.2 (2001): 28-32.
- Malone, Debbie, and Carol Videon. First Year Student Library Instruction Programs. Clip Note #33. Chicago: American Library Association, 2003.
- Marzano, Robert J., Debra Pickering, and Jay McTighe. Assessing Student Outcomes: Performance Assessment Using the Dimensions of Learning Model. Alexandria, Virginia: Association for Supervision and Curriculum Development, 1993.
- Maughan, Patricia Davitt. "Assessing Information Literacy among Undergraduates: A Discussion of the Literature and the University of California-Berkeley Assessment Experience." College and Research Libraries 62.1 (2001): 71-85.



- McBride, Amanda. "Letting Students Shine: Assessment to Promote Student Learning." Focus 31 (1999).
- McDonald, Marilyn, and Gretchen H. Neill. "Standards for Community, Junior, and Technical College Learning Resources Programs." College and Research Libraries News 55 (1994): 572-85.
- Mensching, Teresa B. "Trends in Bibliographic Instruction in the 1980's: A Comparison of Data from Two Surveys." Research Strategies 7.4 (1989): 4-13.
- Mertler, Craig A. "Designing Scoring Rubrics for Your Classroom." Practical Assessment, Research and Evaluation 7.25 (2001).
- Merz, Lawrie H., and Beth L. Mark. Clip Note #32: Assessment in College Library Instruction Programs. Chicago: Association of College and Research Libraries, 2002.
- Miller, Ted K. "Using Professional Standards for Program Assessment and Development". 2002. Net Results. National Association of Student Personnel Administrators. 18 Feb. 2005.  
<<http://www.naspa.org/membership/mem/nr/article.cfm?id=624>>.
- Mittermeyer, Diane, and Diane Quirion. Information Literacy: Study of Incoming First-Year Undergraduates in Quebec: National Library of Quebec, 2003.
- Moskal, Barbara M. "Scoring Rubrics: What, When, and How?" Practical Assessment, Research, and Evaluation 7.3 (2000).
- Moskal, Barbara M., and Jon A. Leydens. "Scoring Rubric Development: Validity and Reliability." Practical Assessment Research and Evaluation 7.10 (2000).
- National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development. The NPEC Sourcebook on Assessment, Volume 1: Definitions and Assessment Methods for Critical Thinking, Problem Solving, and Writing. Harrisonburg, Virginia: National Center for Education Statistics, 2000.
- "NC State Academic Programs GER - Writing, Speaking, and Information Literacy Rationale". June 5 2005.  
<[http://www.ncsu.edu/provost/academic\\_programs/ger/wrtspk/rat.htm](http://www.ncsu.edu/provost/academic_programs/ger/wrtspk/rat.htm)>.
- Nitko, Anthony J. Educational Assessment of Students. Upper Saddle River, New Jersey: Pearson Education, 2004.

---. Educational Assessment of Students. Englewood Cliffs, New Jersey: Prentice Hall, 1996.

Norman, Donald A., and James C. Spohrer. "Learner-Centered Education." Communications of the ACM 39.4 (1996): 24-27.

Nunberg, Geoffrey. "Teaching Students to Swim in the Online Sea." New York Times February 13 2005: 4.

O'Connor, Lisa. "Project SAILS: Clearer Horizons for Information Literacy Assessment." LOEX Quarterly 31.1 (2004): 8-9.

O'Connor, Lisa G., Carolyn J. Radcliff, and Julie A. Gedeon. "Applying Systems Design and Item Response Theory to the Problem of Measuring Information Literacy Skills." College and Research Libraries 63.6 (2002): 528-43.

---. "Assessing Information Literacy Skills: Developing a Standardized Instrument for Institutional and Longitudinal Measurement." Association of College and Research Libraries 10th National Conference. Denver, Colorado, 2001.

Owusu-Ansah, Edward K. "Information Literacy and Higher Education: Placing the Academic Library in the Center of a Comprehensive Solution." Journal of Academic Librarianship 30.1 (2004): 3-16.

Palomba, Catherine A., and Trudy W. Banta. Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education. San Francisco: Jossey-Bass, 1999.

Pausch, Lois M., and Mary Pagliero Popp. "Assessment of Information Literacy: Lessons from the Higher Education Assessment Movement." ACRL National Conference. Nashville, Tennessee: Association of College and Research Libraries, 1997.

Penrose, Ann M. "Race and Gender in the First Year Writing Program." E-mail to Megan Oakleaf. 10 August 2005.

Person, Roland. "Long-Term Evaluation of Bibliographic Instruction: Lasting Encouragement." College and Research Libraries 42.1 (1981): 19-25.

Phillip, Cyndi. "Clear Expectations: Rubrics and Scoring Guides." Knowledge Quest 31.2 (2002): 26-27.

- Pintrich, Paul R. "Motivation: Self-Regulated Learning." Encyclopedia of Education. Ed. James W. Guthrie. Vol. 5. New York: Macmillan Reference USA, 2003. 1698-701.
- . "The Role of Motivation in Promoting and Sustaining Self-Regulated Learning." International Journal of Education Research 31.6 (1999): 459-70.
- Popham, W. James. Test Better, Teach Better: The Instructional Role of Assessment. Alexandria, Virginia: Association for Supervision and Curriculum Development, 2003.
- . "What's Wrong--and What's Right--with Rubrics." Educational Leadership 55.2 (1997): 72-75.
- Prus, Joseph, and Reid Johnson. "A Critical Review of Student Assessment Options." New Directions for Community Colleges 88 (1994): 69-83.
- Ratterau, Oswald M. T. "Information Literacy in Self-Study and Accreditation." Journal of Academic Librarianship 28.6 (2002): 368-75.
- Research and Scholarship Committee of the ACRL Instruction Section. "Research Agenda for Library Instruction and Information Literacy." College and Research Libraries 64.2 (2003): 108-13.
- Resnick, Lauren B. Education and Learning to Think. Washington, D.C.: National Academy Press, 1987.
- Rockman, Ilene F. "Integrating Information Literacy into the Learning Outcomes of Academic Disciplines." College and Research Libraries News 64.9 (2003): 612-15.
- . "Rubrics for Assessing Information Competence in the California State University". 2002. April 10 2005.  
<[http://www.calstate.edu/LS/1\\_rubric.doc](http://www.calstate.edu/LS/1_rubric.doc)>.
- Roth, Lorie. "Educating the Cut-and-Paste Generation." Library Journal 124.18 (1999): 42-44.
- Samson, Sue. "What and When Do They Know? Web-Based Assessment." Reference Services Review 28.4 (2000): 335-42.
- SAS. "Compute Estimates and Tests of Agreement among Multiple Raters". 2006. 5 Jan 2006.  
<<http://support.sas.com/ctx/samples/index.jsp?sid=507&tab=details>>.

- Schmidt, C. James. "Editorial." College and Research Libraries 42.1 (1981): 5.
- Shavelson, Richard J., and Gail P. Baxter. "Linking Assessment with Instruction." A Handbook for Student Performance in an Era of Restructuring. Eds. Robert E. Blum and Judith A. Arter. Alexandria, Virginia: Association for Supervision and Curriculum Development, 1996. IV-7:1 - IV-7:6.
- Shepard, Lorrie A. "The Role of Assessment in a Learning Culture." Educational Researcher 29.7 (2000): 4-14.
- . "Standardized Tests and High-Stakes Assessment." Encyclopedia of Education. Ed. James W. Guthrie. Vol. 7. New York: Macmillan Reference USA, 2003. 2533-37.
- . "Why We Need Better Assessments." A Handbook for Student Performance Assessment in an Era of Restructuring. Eds. Robert E. Blum and Judith A. Arter. Alexandria, Virginia: Association for Supervision and Curriculum Development, 1996. I-2:2 - I-2:7.
- Shonrock, Diana D. Evaluating Library Instruction: Sample Questions, Forms, and Strategies for Practical Use. Chicago: American Library Association, 1996.
- Silver, Edward A. "Performance Assessment." Encyclopedia of Education. Ed. James W. Guthrie. Vol. 1. New York: Macmillan Reference USA, 2003. 134-37.
- Small, Ruth V., B. M. Dodge, and X. Jiang. "Dimensions of Interest and Boredom in Instructional Situations." 1996 National Convention of the Association for Educational Communications and Technology. Indianapolis, Indiana, 1996.
- Small, Ruth V., Nasriah Zakaria, and Houria El-Figuigui. "Motivational Aspects of Information Literacy Skills Instruction in Community College Libraries." College and Research Libraries 65.2 (2004): 96-121.
- Smalley, Topsy N. "Bay Area Community Colleges Information Competency Assessment Project". 2003. November 17 2003.  
<<http://www.topsy.org/ICAP/ICAPProject.html>>.
- Smith, Kenneth R. New Roles and Responsibilities for the University Library: Advancing Student Learning through Outcomes Assessment: Association of Research Libraries, 2000.
- Smith, Susan Sharpless. Web-Based Instruction: A Guide for Libraries. Chicago: American Library Association, 2001.

- Snavey, Loanne L., and Carol A. Wright. "Research Portfolio Use in Undergraduate Honors Education: Assessment Tool and Model for Future Work." Journal of Academic Librarianship 29.5 (2003): 298-303.
- Spady, William G. "Outcome Based Education." Encyclopedia of Education. Ed. James W. Guthrie. Vol. 5. New York: Macmillian Reference USA, 2003. 1827-31.
- Stemler, Steven E. "A Comparison of Consensus, Consistency, and Measurement Approaches to Estimating Interrater Reliability." Practical Assessment, Research, and Evaluation 9.4 (2004).
- Stevens, Dannelle D., and Antonia Levi. Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback, and Promote Student Learning. Sterling, Virginia: Stylus, 2005.
- Stiggins, Richard J. "The Development of Performance Assessment Exercises." A Handbook for Student Performance in an Era of Restructuring. Eds. Robert E. Blum and Judith A. Arter. Alexandria, Virginia: Association for Supervision and Curriculum Development, 1996. V-8:1 - V-8:9.
- Sweet, David. "Performance Assessment." Education Research Consumer Guide 2 (1993).
- Thompson, Gary B. "Information Literacy Accreditation Mandates: What They Mean for Faculty and Librarians." Library Trends 51.2 (2002): 218-41.
- Tierney, Robin, and Marielle Simon. "What's Still Wrong with Rubrics: Focusing on the Consistency of Performance Criteria across Scale Levels." Practical Assessment, Research, and Evaluation 9.2 (2004).
- Ursin, Lara, Elizabeth Blakesley Lindsay, and Corey M. Johnson. "Assessing Library Instruction in the Freshman Seminar: A Citation Analysis Study." Reference Services Review 32.3 (2004): 284-92.
- Waller, Consuela Askew, et al. "Two (or Twelve) Heads Are Better Than One: Addressing the Complexities of Working with ACRL's Information Literacy Outcomes to Develop Assessment Tools." Association of Research Libraries 11th National Conference. Charlotte, North Carolina, 2003.
- Warmkessel, Marjorie M. "Assessing Abilities of Freshmen to Reconcile New Knowledge with Prior Knowledge." Assessing Student Learning Outcomes for Information Literacy Instruction in Academic Institutions. Ed. Elizabeth Fuseler Avery. Chicago: Association of College and Research Libraries, 2003. 249-56.

- Warner, Dorothy Anne. "Programmatic Assessment: Turning Process into Practice by Teaching for Learning." Journal of Academic Librarianship 29.3 (2003): 169-76.
- Werking, Richard Hume. "Evaluating Bibliographic Education: A Review and Critique." Library Trends 29.1 (1980): 153-72.
- Wiggins, Grant. "Creating Tests Worth Taking." A Handbook for Student Performance in an Era of Restructuring. Eds. Robert E. Blum and Judith A. Arter. Alexandria, Virginia: Association for Supervision and Curriculum Development, 1996. V-6:1 - V-6:9.
- . Educative Assessment: Designing Assessments to Inform and Improve Student Performance. San Francisco: Jossey-Bass, 1998.
- Williams, Janet L. "Creativity in Assessment of Library Instruction." Reference Services Review 28.4 (2000): 323-34.
- Wolfe, Edward W., Chi-Wen Kao, and Michael Ranney. "Cognitive Differences in Proficient and Nonproficient Essay Scorers." Written Communication 15.4 (1998): 465.
- Wolff, Ralph A. "Using the Accreditation Process to Transform the Mission of the Library." New Directions for Higher Education 90 (1995): 77-91.
- Young, Jeffrey R. "Testing Service to Unveil an Assessment of Computer and Information Literacy." Chronicle of Higher Education November 12 2004.
- Young, Rosemary, M., and Stephena Harmony. Working with Faculty to Design Undergraduate Information Literacy Programs. How-to-Do-It Manuals for Librarians. New York: Neal-Schuman, 1999.
- Young, Virginia E., and Linda G. Ackerson. "Evaluation of Student Research Paper Bibliographies: Refining Evaluation Criteria." Research Strategies 13.2 (1995): 80-93.
- Zeller Jr., Tom. "Measuring Literacy in a World Gone Digital." New York Times January 17 2005.