



Learning by Doing: Developing a Baseline Information Literacy Assessment

Stephen “Mike” Kiel, Natalie Burclaff, and Catherine Johnson

abstract: This paper details the design and implementation of an initial baseline assessment of information literacy skills at the University of Baltimore in Maryland. To provide practical advice and experience for a novice audience, the authors discuss how they approached the design and implementation of the study through the use of a rubric-based authentic assessment, employing a pretest and posttest delivered through a course management system. They also present lessons learned through the process of assessment focused on norming, test design and delivery, and the importance of institutional support and flexibility.

Introduction

There has been increasing interest in assessment during recent years in higher education as well as in libraries specifically. According to Christina Leimer, the 2009 National Institute for Learning Outcomes Assessment (NILOA) survey revealed that 60 percent of provosts believe they need “more faculty engagement and more technical expertise to strengthen the assessment of learning on their campuses.”¹ The Association of American Colleges & Universities (AAC&U) has been instrumental in meeting this need through the development of VALUE (Valid Assessment of Learning in Undergraduate Education) rubrics and by providing numerous conference workshops. Libraries and their professional organizations have also attempted to fill this gap in assessment expertise. Megan Oakleaf’s *The Value of Academic Libraries: A Comprehensive Research Review and Report* prepared for the Association of College and Research Libraries (ACRL) identified areas of research for libraries, emphasizing evidence-based practices

portal: Libraries and the Academy, Vol. 15, No. 4 (2015), pp. 747–766.

Copyright © 2015 by Johns Hopkins University Press, Baltimore, MD 21218.



to demonstrate how libraries impact university values.² As a result of the report, ACRL created the Assessment in Action (AiA) program to provide training and an opportunity for librarians to get firsthand experience with assessment of student learning.

In 2013 at the University of Baltimore in Maryland, a number of factors indicated that the time was right for librarians to engage in a more intense evaluation of information literacy instruction. At that time, a new associate provost was hired to focus on engaging and coordinating campus-wide assessment. Part of the provost's role is to coordinate the submission of a self-study in the spring of 2017 to the Middle States Commission on Higher Education (MSCHE), the accrediting body of the University of Baltimore. This self-study must demonstrate the university's compliance with fourteen MSCHE standards. Two standards directly reference assessment, while the remaining twelve indicate that the university's continuous improvement should be evidence-based. This upcoming self-study comes after a 2012 review in which MSCHE provided recommendations that the university should enhance and institutionalize assessment of student learning outcomes in courses and programs of study, including general education.

Around the same time, a new director of the Writing Program, who had a desire to improve programmatic assessment, joined the faculty and, in the 2012 academic year, the university revamped its general education program with a new focus on building skills and competencies across a student's education. Information literacy (IL) was identified as a component of new university learning goals, and a new mandatory sophomore seminar included IL skills as one of its components. In addition, library faculty had been striving to ensure a scaffolded delivery, first introducing general skills, then reinforcing these and introducing advanced and discipline-specific IL skills later in the curriculum. The librarians worked with individual academic programs and college administrators to fully implement a scaffolded instruction program. The combination of these factors made it the right time to take advantage of the AiA program to create a campus team to evaluate information literacy skills at multiple points in the curriculum. An additional motivating factor was that the AiA program provided the opportunity to develop expertise in assessment, which no University of Baltimore librarians possessed when the project began.

A recent study found that about 57 percent of all library periodicals include research content.³ However, most of these papers focus on the results and not the methods, and the majority use surveys as the primary method of data collection.⁴ For librarians beginning to learn assessment techniques, most research articles offer little insight as to how

Many librarians enter the field with little or no formal training in evaluation of skills or research methods.

the authors approached the decisions involved in designing their studies, challenges with implementation, and factors that contributed to a successful assessment plan.⁵ This lack can create unrealistic or vague models of assessment for novices who hope to learn from or replicate assessment studies. Additionally, only slightly more than half of library and information science programs require a research

methods course,⁶ which can provide the foundation for understanding assessment strategies and techniques. Consuella Askew and Eileen Theodore-Shusta found that assessment was referenced in only 10 percent of library and information science course



descriptions,⁷ which implies that many librarians enter the field with little or no formal training in evaluation of skills or research methods. The goals of assessment and research differ,⁸ but the lack of training in either can leave library professionals unprepared or unsure how to engage in these processes. Therefore, it is unsurprising that few articles focus on the assessment planning and implementation process.

The following article will discuss how librarians with novice assessment skills at the University of Baltimore approached the design and implementation of a large-scale baseline study of information literacy skills. The authors will also highlight lessons learned from the experience. The purpose of this discussion is to encourage other academic librarians to begin the assessment process and to learn from the successes and setbacks inherent in identification of targets, planning, and implementation of the first round of a multiyear assessment project.

Identification of Goals, Partners, and Targets

The first step in any assessment project is the identification of goals and targets. The Langsdale Library at the University of Baltimore had, in 2012, created an assessment committee. This committee consists of the three authors of this study, who are two faculty librarians and the head of information literacy initiatives. Before embarking on the project, the committee had to identify broad goals, find collaborators to help the assessment run more smoothly, and target points in the curriculum where the evaluation of skills would take place.

Setting Goals

The assessment committee began by surveying the current information literacy and assessment landscape at the university. Through this process, the committee discovered that past appraisals had been sporadic, irregular, and rarely used to implement meaningful change. In fact, little was known about students' IL competencies. As a result, the committee decided to conduct a baseline study of students' information literacy skills. A baseline study provides "a known level of student achievement against which new measurements can be compared."⁹ Baseline studies are an experimentally sound method for instructors to use to test factors that improve student learning,¹⁰ but they make up only a small amount of methodologies used in librarian research.¹¹

Baseline studies are an experimentally sound method for instructors to use to test factors that improve student learning, but they make up only a small amount of methodologies used in librarian research.

The assessment committee selected the methodology of a baseline study for a number of reasons. First, a baseline study would provide a reliable method to identify problem areas in students' information literacy competencies.¹² Second, a baseline study exposes benefits and drawbacks of various pedagogies and styles of library sessions. It also reveals, if the study is repeated, the impact of information literacy sessions on student



learning and success. This methodology, therefore, has the potential to tell us not only about student learning but also, over time, about the strengths and weaknesses in librarians' teaching. Third, a baseline study can provide data that can be disseminated to faculty to offer a fuller understanding of student information literacy skills.¹³ A baseline study therefore seemed necessary to give faculty an understanding of students' current skills in order to gain support for the library's work and potential revisions to the IL program based on assessment results. The baseline study also allows for subsequent assessment efforts based on the results and can be repeated periodically to track changes over time.

Oakleaf's *The Value of Academic Libraries*, driven by an increasing national conversation focused on measuring student learning and institutional effectiveness, calls on

After deciding on a baseline study, the library assessment committee knew that, to be most successful, they needed to collaborate with a variety of campus partners.

librarians to demonstrate "the value of academic libraries in clear, measurable ways."¹⁴ Librarians and administrators at the University of Baltimore hoped to learn more about the myriad ways the library contributes to student success. After deciding on a baseline study, the library assessment committee knew that, to be most successful, they needed to collaborate with a variety of campus partners. Working with partners would not only improve the accuracy and effectiveness of the assessment but also position the library to become a more

valuable participant in student learning.¹⁵ Prior to engaging in project design specifics, the authors sought natural collaborators across the university who could bring unique strengths to the project.

Identifying Partners

The three members of the library's assessment committee identified four additional campus partners. Table 1 details the strengths that these partners, and the authors, brought to the project. Some job titles have been modified to strip away institutional language and clarify their role in the university. Each was selected because of the knowledge, expertise, and resources he or she could provide to ensure successful project completion.

The librarians recognized the need for additional assessment expertise that could be provided by the associate provost for assessment and the director of the Bank of America Center for Excellence in Learning, Teaching, and Technology at the university. Both individuals were recent additions to the University of Baltimore community, hired in part to enhance the university's assessment practices in preparation for the upcoming accreditation review. In addition to adding expertise in rating student skills, the librarians selected team members who could serve as natural collaborators in the project. The writing director could provide access to data in a range of courses that might otherwise not be available, and the head of instructional technologies could furnish admittance to the course management system, the software that the university used both to create and distribute course content and track student performance. The head of instructional technologies could also provide recommendations on the online implementation and delivery of the assessment. Together with the three authors, who made up the library



Table 1.
Positions and strengths of campus team members

Role	Abilities
Faculty librarians	Knowledge of library assessment practices
Head of information literacy initiatives	Understanding of information literacy program goals and structure
Associate provost for assessment	Assessment expertise and the ability to promote the project, secure funding, and connect the team to others on campus who could be a resource; navigation of the campus political culture
Director of University Writing Program	Buy-in and investment in the success of the project from the Writing Program, faculty compliance
Head of instructional technologies	Understanding of digital campus resources and data acquisition and management, interfacing with online learning platform
Director of Bank of America Center for Excellence in Learning, Teaching, and Technology	Expertise in assessment, connection to experts in the field and deep knowledge of similar studies

assessment committee, these individuals formed the campus team that completed the AiA project.

Librarians on the assessment committee identified and approached each team member to explain the project, goals, and timeline. In doing so, the committee emphasized professional benefits, such as opportunities for publication, presentation, and training. Such an explanation of benefits is a good practice to encourage a strong investment in the project from team members outside of the library, especially in a multiyear project.¹⁶ Though working on this project involved a significant time commitment, participants were motivated by the opportunity to build a campus model of assessment that could be replicated. A modest grant from the University of Baltimore Foundation enabled the authors to invite team members involved in the assessment project to present results at the 2014 American Library Association Annual Conference.

Targeting the Assessment

Once assembled, the campus team agreed that, given the size of the team and the motivation driving the project, the group should aim to collect as much meaningful



data as possible. Megan Oakleaf and Neal Kaske suggest that one means of collecting significant, reliable data is to gather results from as large a student population as feasible.¹⁷ The team also made a strategic choice to assess the populations that consistently receive information literacy education from librarians. Using this model, the team could potentially draw conclusions about the effectiveness of the current IL efforts as well as have meaningful places to implement changes based on the results. The campus team identified two three-credit courses as the targets for assessment, INFO 110: Introduction to Information Literacy, and WRIT 300: Composition and Research. INFO 110 is taught by librarians and designed for first- and second-year students. It is required for any student entering the university with fewer than 45 credits. WRIT 300 is a required course for all students, including the university's large transfer population. Each section of WRIT 300 has an embedded librarian, and, although the level of library involvement varies from instructor to instructor, a library instruction session is usually included at some point during the semester. A baseline evaluation in these courses positioned the team to capture the information literacy skills of students from a variety of educational backgrounds who have had a variety of interactions with librarians at the University of Baltimore. While assessment in the identified courses would provide a snapshot of IL skills of students only at the first-year and junior level, the team attempted to craft a structure for assessment that allowed for expansion to other points in the curriculum in the future.

Initial Development of Study Instruments

Having assembled a team to support their efforts and identified target populations and courses to study, the authors turned their attention to the development of the methods to conduct the assessment. In light of the library's long-term goals for a more intentional and scaffolded information literacy program, the team wanted to get as full a picture of current student skills as possible in order to accurately target development of the program in the future. After reviewing the IL program learning outcomes, the team identified two methods of baseline data collection. Outcomes related to the application of skills would be measured through the use of a rubric-based assessment of student work, whereas outcomes focused on knowledge and comprehension would be measured using a multiple-choice or short-answer test.

Designing the Study

Much of the current research conducted by librarians uses only one assessment technique.¹⁸ Brandy Whitlock and Julie Nanavati outline the strengths and weaknesses of various approaches, illustrating the advantages of a two-pronged approach.¹⁹ Using authentic artifact-based assessment alone would make it impossible to demonstrate a change as a result of instruction during the course of the semester. A pretest and posttest ensured that improvement could be tracked. The authentic assessment complemented the pretest and posttest by asking students to perform real-world tasks requiring meaningful application of their skills, thus enabling a more sophisticated and natural test of student abilities. The pretests and posttests reveal whether students know information-



literacy-related concepts, whereas the authentic assessment reveals whether and how well students demonstrate information literacy skills in their work. Additionally, some of the IL learning outcomes could be evaluated using both methods, allowing for a comparison of student performance in multiple modalities.

Interest in authentic assessment has been driven by a desire for a better understanding of student application of information literacy skills. Authentic assessment is defined in a variety of ways through the literature.²⁰ For the purpose of this project, the campus team chose to define authentic assessment

as a school-based project in which students were expected to conduct research and integrate sources into their work in a meaningful way. The authors acknowledge that while this work is authentic to the educational environment, this task is not necessarily indicative of how students' information literacy skills might manifest themselves in other environments (such as work-related tasks, community projects, and the like). However, even if imperfect,

authentic assessments have great value because they are often more valid than other appraisal methods to students and external evaluators.²¹ Due to the time involved for an authentic assessment of multiple sections of two courses, final research assignments were determined to be the ideal artifact because they would represent the height of student skills. The team rated these assignments using rubric-based scoring.

The authentic assessment complemented the pretest and posttest by asking students to perform real-world tasks requiring meaningful application of their skills, thus enabling a more sophisticated and natural test of student abilities.

Building the Rubric

The rubric used in the baseline study to score the student artifacts was developed through a lengthy process of norming and revision. Initially, the assessment committee began to create the rubric by adapting the university library's information literacy learning goals. Since the rubric was intended for assessment of student work at multiple points in the curriculum, the authors attempted to design a rubric reflecting the entire range of student information literacy skills from a novice to a student who has mastered IL skills at the undergraduate level. The library assessment committee also worked to develop a rubric that could evaluate as many as possible of the ACRL Information Literacy Competency Standards for Higher Education and the performance indicators for each standard to ensure that the baseline generated by the assessment told a complete story about student IL skills. At the time the rubric was developed, work had not yet progressed far enough to map the rubric to ACRL's new Framework for Information Literacy for Higher Education, but similar processes can be used in the future to create rubrics using the Framework as a foundation. The final rubric had five dimensions, each with a guiding question to assist raters, mapped to the Standards and indicators as presented in Table 2.

The committee modeled its rubric in part on the AAC&U VALUE rubrics. Drawing on the experience of faculty developing the VALUE rubrics, the committee designed a



Table 2.
Rubric mapping to information literacy standards

Rubric category	Rubric guiding question	ACRL Standard
Position	How clearly has the writer identified his or her position?	Standard 1: Indicators 1, 4
Relevance	How relevant are the sources?	Standard 3: Indicators 2, 4
Use	How well did the writer use the information?	Standard 3: Indicators 1, 3, 4
Style	How well did the writer cite the source material?	Standard 5: Indicator 3
Seamlessness	How effective is the integration of source material?	Standard 4: Indicators 1, 3

rubric with four levels of performance, with an additional option (“0”) for work that did not meet minimum levels of performance. In the case of the VALUE rubrics, four levels provided sufficient detail to faculty teams to differentiate between student performances at the institutional or program level without being so fine-grained as to hamper the ability to reach broad consensus about student achievement.²² While designing the rubric, the authors followed the process described in *Introduction to Rubrics* by Dannelle Stevens and Antonia Levi. They first determined for each category the attributes that a student would display at the highest or “mastery” level followed by those at the lowest or “developing” level. Establishing the two extremes allowed the authors to more effectively make the difficult and more nuanced decisions required to fill in the middle levels.²³ Defining rubric levels in this way also helped the committee view each level as part of a continuum of skills, rather than a measure of what students would be expected to do in a given year of their degree program. In the development process, the committee used positive language that expressed the presence of a skill rather than its absence to define each set of included criteria. To use the “Position” metric as an example, work at the most novice end of the rubric, “Developing,” was defined as “Thesis or research question present, but parts are too broad or poorly defined,” whereas work at the highest level of the rubric, characterized as “Mastery,” is defined as “Effectively defines the appropriate scope of the research question or thesis and situates it within a larger conversation surrounding the topic.” In both cases, the language is positive and defines what the students achieve in their papers, not what they fail to do. Using positive language prevents rubric scorers from characterizing papers that appear on lower levels of the rubric as failures and those on the higher levels as successes.²⁴



Initially, development of this rubric involved only the library assessment committee. Individual librarians were assigned different dimensions to describe and then, as a group, the committee debated and refined the language over the course of several meetings. Once a working draft was developed, the authors attempted to pilot the rubric amongst themselves on a small sample of three papers, which led to some important revisions. Pilot testing is a key step in the development process that can help researchers identify areas of survey and study instruments that contain errors

or could benefit from redesign.²⁵ Chief among the changes made after pilot testing was more detailed language describing student behaviors and evidence at each level, and the addition of a fifth dimension for evaluation. The earliest draft of the rubric lacked the aforementioned “Position” category, which the authors realized would be important to capture separately when the rubric was tested. The pilot testing process was repeated after the initial round of revisions and, having achieved agreement, the authors deemed it appropriate to move on to testing the rubric outside the library.

Using positive language prevents rubric scorers from characterizing papers that appear on lower levels of the rubric as failures and those on the higher levels as successes.

Developing the Pretest and Posttest

Parallel to the rubric development process, the authors developed a pretest and posttest as a means to assess competencies reflecting as many of the ACRL Standards as possible. A key decision in this process was the attempt to base incorrect answers on actual student responses and behaviors previously observed. As an example, one question focused on the quality and appropriateness of research questions:

If you were assigned a 10-page research paper about drug treatment programs, the best research question would be:

- a. Where are methadone centers located in the United States?
- b. What are the most effective methods for drug treatment of teen mothers used at the Allegany County Addictions Services Joseph S. Massie Unit?
- c. What barriers exist when trying to implement an effective drug treatment program in the prison system?
- d. What factors contribute to illicit drug abuse among teenagers?

Here the correct answer is “c.” The other answers reflected a number of frequent student problems observed at the University of Baltimore. A student who answers “a” most likely misunderstands the nature of college level work and research. Response “b” indicates a student defining a need for information that may be proprietary and unavailable. A student who answers “d” appears to be on the right track but lacks the correct scope or specificity and does not adequately address the topic assigned.

Unlike the rubric, which could be tested by its authors directly, the pretest and posttest were checked for clarity and comprehension by administering them to library student assistants. As a result of the assistants’ employment, they could be reasonably



expected to have sufficient familiarity with the concepts and terminology of the tests. This helped the team refine the test, resulting in thirteen questions testing knowledge of searching mechanics and strategies, citation, and comprehension and evaluation of sources. In an effort to help the assessment committee make more meaningful comparisons of student performance, the test also asked students for demographic information as well as their previous exposure to library instruction.

Implementation

To fully implement the baseline study that had been designed by the assessment committee

Implementation is a challenging phase in many studies because researchers want to follow best practices to produce the most valid results. However, on many occasions researchers must deviate from their initial plans due to local conditions.

required development and execution of a data-collection plan, the scoring of student work using the rubric, and subsequent analysis of results. The authors developed a vision for how this process would flow but had to adapt and revise the plan as the project progressed. Implementation is a challenging phase in many studies because researchers want to follow best practices to produce the most valid results. However, on many occasions researchers must deviate from

their initial plans due to local conditions.²⁶ The following section will highlight some local conditions and describe how the project was implemented as a result.

Preparing for Data Collection

To ensure compliance with ethical researching practices, the authors first submitted the research plan to the university's Institutional Review Board (IRB). The IRB required a plan to give the participants a consent form, which was created and attached to the pretest and posttests. This requirement meant an authentic assessment of coursework could only be collected from students who consented either through the pretest or posttest. Data collection could not be anonymous because the authors needed to track students at multiple points in one semester, through the pretest, posttest, and research paper, and through multiple points in their college career, in the hope that the initial baseline study would become the foundation of a multiyear evaluation. However, the data would be kept confidential, as only the three librarians from the assessment committee would have access to student information.

The initial round of data collection began in the fall semester of 2013. To assuage concerns and explain the collection methods, the head of information literacy initiatives met with the upper-division writing instructors shortly before the semester started. At that meeting, she described the reasons behind collecting data about information literacy skills at multiple points in the curriculum and why the upper-division writing course had been selected. One of the main concerns from teaching faculty was that the university would view these assessments as a reflection on their teaching. It became



important to emphasize that this investigation was a baseline study looking at skills that these instructors do not explicitly teach, and a way for librarians to evaluate if different approaches to library instruction were necessary.

Collecting Data

Collecting the pretest and posttest responses was straightforward because they were embedded into the course management system. The program manager for e-learning pulled test results from the courses the team identified and exported them to a spreadsheet. The team then collected papers from those students who gave consent for participation in the study through the pretest or posttest. However, the collection of papers was not without challenges. For example, some final research assignments were group projects, which could not be used because the scores could not be connected to a specific pretest or posttest result. Additionally, some faculty defined the research paper loosely and asked students to write about their research process rather than incorporate their research into a final product. Because these assignments did not align with the rubric, papers from those faculty members' sections were excluded. In some cases, where the student work was not submitted electronically, the campus team had to make an extra effort to locate, collect, and copy paper assignments. The team collected original copies of the papers, removed all identifying information, and assigned each a random number. The head of information literacy initiatives tracked the random numbers and withheld them from the rest of the raters. This method allows raters to mitigate halo bias, the potentially inaccurate judgment that occurs when educators grade based on their previous experiences with students and not qualities of the work itself.²⁷

One of the main concerns from teaching faculty was that the university would view these assessments as a reflection on their teaching. It became important to emphasize that this investigation was a baseline study looking at skills that these instructors do not explicitly teach.

Scoring

The next challenge was coordinating the rubric scoring of the 212 papers that had been collected. After initial rubric creation by the assessment committee, the entire campus team began a broader norming and scoring process. Norming is an important step when using rubrics to evaluate student work. Norming ensures that the rubric is as strong a tool as possible because the process is designed to ensure consistency and reliability in scoring.²⁸ When scoring of the papers began in spring 2014, the task was divided between the various campus team members with the addition of some writing professors who were recruited. In the first round,

Norming ensures that the rubric is as strong a tool as possible because the process is designed to ensure consistency and reliability in scoring.



each person read eight papers, then came back to large group to discuss how they had scored individual artifacts. Each paper was assigned to at least two scorers to facilitate conversations about discrepancies in scores. Adjacent agreement was used for inter-rater reliability, meaning the scores needed to be within one point to constitute consensus.²⁹ The team also used adjacent agreement as a signal that raters were scoring consistently and were ready to stop norming and start the full scoring. The head of information literacy initiatives distributed the anonymous papers to the raters electronically, so that each paper was read twice, and no rater received a paper written by a student enrolled in his or her class. The raters had three weeks to complete their ratings and submit their scores via a Google form created for this purpose. The team then compiled the ratings in a spreadsheet and interpreted the baseline data using SPSS, a software package for statistical analysis. In addition to each set of student scores resulting from the rubric, data from the pretest and posttest were exported from the e-learning system and included in SPSS.

Lessons Learned

While not the focus of this paper, analysis of the initial data highlighted particular information literacy skills that students struggled with at both the freshman and junior levels. The baseline study thus provided useful information, which has informed the librarians' teaching strategies in both courses. As part of the larger assessment project, the baseline study was only the first stage. The data-collection process was repeated again in the fall of 2014, and data collection and analysis have continued. However, the experience of planning and implementing a large-scale project taught the authors a great deal about assessment. In particular, the team learned useful lessons regarding norming and scoring, the limitations of testing, and the importance of strong campus partnerships, which will benefit them as they move forward in the completion of future rounds of this assessment and design future studies.

Norming and Scoring with the Rubric

One member of the team had much experience working with students enrolled in developmental courses. This led him to score the student work more generously than another member who often teaches graduate students and consequently gave far lower scores.

The process of norming revealed a number of unexpected issues. One of these was the nature of the artifacts that were evaluated. There was little control or standardization in the assignments in either of the courses that were examined. The majority of the artifacts came from WRIT 300 because the course had a larger enrollment than INFO 110. Various sections of that course included the following final assignments: lengthy grant proposals, traditional research papers, and short reflections on poetry. This diversity was often a challenge because students could use sources in a wide variety of ways in the assignments, making a direct comparison of the rubric scores difficult. Ultimately, the group was



only able to finish norming and scoring by discarding some artifacts. In the future, the librarians will need to work on developing research-based assignments with WRIT 300 faculty to ensure a greater consistency in outcomes.

The assessment process also revealed the different ways in which faculty outside of the library understood and evaluated information literacy concepts compared to librarians, which affected their rubric scoring. The strongest example of this occurred when attempting to rate the relevance of sources students used to their topic. Library faculty almost universally proceeded by examining the bibliography and comparing it to the stated purposes of the paper. In contrast, other faculty tended to carefully examine which individual quotations or other information were extracted from the sources and how students used this information before determining if the source was relevant, mostly ignoring the list of references. This difference resulted in divergent rubric scores for source relevance, and raters engaged in long discussions during the initial norming session. Additionally, the populations of students they had previously taught colored faculty perceptions and expectations. One member of the team had much experience working with students enrolled in developmental courses. This led him to score the student work more generously than another member who often teaches graduate students and consequently gave far lower scores. Both faculty members, despite anonymized papers, still exhibited biases the authors had attempted to mitigate in their initial project design. These unexpected and dramatic differences in interpretation meant that norming was a far more time-consuming and intense process than initially expected.

In an attempt to navigate rater disagreements, the team followed several of the rules outlined in Claire Holmes and Megan Oakleaf’s 2013 paper “The Official (and Unofficial) Rules for Norming Rubrics Successfully.”³⁰ Modeling a practice learned in an AAC&U training session at the University of Baltimore, the team recorded scores of all raters on a whiteboard with unique colors corresponding to individual raters. This allowed for two patterns to be easily recognized, the pattern of an individual rater’s score and the overall distribution of scores. This visualization technique enabled the group to concentrate on areas of significant disagreement for further discussion. After an initial agreement had been reached, pairs of individual divergent raters discussed their ratings in subsequent norming sessions and reported any concerns to the rest of the group. One helpful process to achieve greater inter-rater consensus was articulating which changes between rubric levels were of the greatest importance. The team decided that the biggest distinction would be between low-scoring work (1 and 2 on the rubric) and high-scoring work (levels 3 and 4). This focused conversations on any differences between a level 2 score and a level 3 score.

Another area in which the authors learned valuable lessons was in determining the range of scores raters were permitted to assign. Initially, the committee allowed raters to give half scores, in effect acknowledging that a student exhibited exceptional work at a particular level but had not met the necessary threshold to move to the next level on

Forcing the team to use only whole numbers forced raters to engage more meaningfully with the difficult questions of what “developing” through “mastery” level work looked like in practice.



the rubric. This method made scoring agreement challenging. It also called into question the design of the rubric itself. If the language was not discrete enough between scores to reflect various levels of students, then the rubric was not a useful instrument for this assessment. Forcing the team to use only whole numbers led the team to refine the rubric to make it clearer and forced raters to engage more meaningfully with the difficult questions of what “developing” through “mastery” level work looked like in practice. For similar reasons, the authors decided that raters would not be allowed to assign a score of “not applicable.” This omission necessitated that raters grapple with the wide variety of assignments and the various ways research is used across assignments. Both of these decisions reflected the team’s decision to treat all of the research-based assignments equally in order to give the assessment more validity or value as an impetus for curricular reform. As a result of the challenging norming and scoring sessions, the campus team recognized the need for additional campus-wide education on the design, revision, and use of rubric-based assessments.

Test Limitations

While the experience of norming with the rubric was the most educational, the team also learned much from the process of administering the pretest and posttests. According to the authors’ plan, data from the test and rubric would complement one another and allow the assessment team to get a full picture of what a student knows and does. In reality, what the team could learn from the test was somewhat limited. A major factor that contributed to this limitation was the difficulty in getting participants. Despite promotion of the pretest and posttest to students and instructors via face-to-face and e-mail communication, only 12 percent of all students completed both the pretest and posttest. This small sample severely limited the authors’ ability to draw meaningful conclusions about improvements during the semester, and between students at the freshman and junior level. The timing of the posttest, which coincided with the end of the semester, and the relative invisibility of the test in the course management system likely factored

Students will more likely respond if they see a survey as a way to facilitate change by providing feedback and if they trust the institution will actually listen to the responses.

into the low participation rate. A 2012 study on undergraduate participation in surveys reinforced findings from earlier survey participation research, which could improve future studies. Students will more likely respond if they see a survey as a way to facilitate change by providing feedback and if they trust the institution will actually listen to the responses.³¹ Despite having this information written on the electronic consent form, face-to-face communication of how the authors will use the survey information

might result in an increased response rate. Additionally, survey fatigue has become common among undergraduates. They often feel that their reward, whether an actual giveaway or an environmental change, is not worth the cost, in time and energy, of taking the survey.³² Making the rewards clear and decreasing the amount of time investment could improve the response rate in the future. A well-timed and well-designed e-mail

follow-up may also help improve participation rates in future rounds of assessment.³³

Beyond difficulties with ensuring participation, the test had limitations. This was the first time this instrument had been used. When designing the test, the authors followed best practices of keeping the test short, avoiding negative wording, varying the position of correct answers, and avoiding answers of unequal lengths.³⁴ However, more time should have been spent ensuring the validity of the test, similar to the time invested norming the rubric. To validate the test, the authors did rounds of content validity with other librarians to evaluate the appropriateness of the questions. Additionally, as previously noted, they conducted a pilot test by having students outside of the sample take the test. However, they did not rigorously review the responses with those students to discover why they selected the answers they did. More diligent pilot testing might have confirmed that the students’ reasoning was similar to what the authors intended with the questions and highlight any confusingly written questions or multiple-choice options.³⁵

Importance of and Takeaways for Campus Partners

Initial enthusiasm for the baseline assessment ran high, but enthusiasm alone might not sustain buy-in over a multiyear project. Project leaders need to regularly report out the value of the assessment to stakeholders and to communicate expectations to the assessment team.³⁶ To help build buy-in, the authors had regular conversations with academic departments to continue awareness of the project and used faculty recommendations to bring in a new team member with the appropriate skills needed. Having the team comprised of members from different units on campus enabled the library to get campus-wide support for the project. Additionally, the associate provost came from an assessment background and understood the value of a baseline study that would extend to a multiyear project. Continued conversations with the associate provost allowed for a wider dissemination of the library-based study. When the provost spoke with other divisions on campus about assessment in preparation for the upcoming accreditation review by the MSCHE, the information literacy baseline study was often mentioned.

Keeping library administration and supervisors involved in the process and partnering with other faculty on campus can ensure that any assessment project will have increased value to groups beyond the individuals performing the assessment. At the University of Baltimore, project partners outside the library found the assessment useful because reading large quantities and varieties of student work gave them insights into student assignments outside of the specific skills being assessed. The study provided faculty a chance to reflect on what might be helpful to assess in their own disciplines and assignment design. The project was also a lesson in the difficulty of assessing student work outside one’s own area of expertise. Simply because someone has done rubric-based assessment before is no guarantee that

Keeping library administration and supervisors involved in the process and partnering with other faculty on campus can ensure that any assessment project will have increased value to groups beyond the individuals performing the assessment.



the individual can smoothly pivot to evaluate other kinds of evidence. This has important implications for the assessment of campus-wide initiatives such as general education, because no individual will be equipped to assess every student learning outcome. It is therefore important to build in sufficient time for norming and training, especially in institutions where faculty may not be used to rubric-based assessment.

Librarians in supervisory positions can also derive some benefits from staff participation in a project of this type. Primarily, the study can be used in a variety of ways to demonstrate the value of the library and of librarians' work to university administration. In addition to communicating the importance and nature of information literacy outside of the library, the project and the methods learned through its execution can be applied to the assessment of other library services. The study also highlights the value of collaboration between faculty and librarians in the creation of assignments and the value of the library as a connection point between departments.

Finally, one contributor to the success of the assessment project was the ability of each team member to adapt to unexpected changes. This was especially true in this

The study also highlights the value of collaboration between faculty and librarians in the creation of assignments and the value of the library as a connection point between departments.

case because natural turnover in the campus team happened over the first year and a half of the project. Over the course of the project's initial conception and completion, many key personnel and leaders at the university were replaced, including half of the original team. When one person left, the library assessment committee worked to find a new team member who could bring similar strengths to the project. However, there could be no assurance

that the new hires or those in related positions would be equally excited about the project. One of these positions was the key post of writing director, who would be important to implementation and compliance. Additional support was needed beyond the team, as well, to ensure the project's success. Because of the librarians' inexperience with data software, the team sought help from colleagues. A psychology professor with close ties to the library volunteered her time to show a member of the team how to analyze data using SPSS. The psychology professor visited multiple times and answered any follow-up questions related to data analysis. Without her assistance, the librarians would have needed to obtain additional training. Future university assessment projects will seek to build on the strengths of this project by engaging partners who bring expertise, a network of colleagues, and enthusiasm to the project.

Continuing the Assessment Cycle

Using assessment results to make changes to instruction essentially "closes the loop" in the cycle of assessment of information literacy instruction.³⁷ After enacting decisions, librarians then move to reviewing new learning goals, starting the cycle over.³⁸ Multiyear assessments provide ongoing opportunities to improve by making data-informed adjustments in each loop.³⁹ Despite some of the limitations and difficulties outlined earlier, the librarians have made a number of changes to their instruction and the curriculum as a



result of this first round of assessment. In INFO 110 instruction, librarians have begun the process of altering the learning objectives for the course. The current draft decreases the number of goals to focus on the foundational skills that the library recognizes as essential for students before they get to their majors.

WRIT 300 instruction has undergone more substantial changes. After assessment results revealed WRIT 300 students often struggle to integrate sources in their work, the committee developed a uniform model for instruction across all sections. By sharing results with the Writing Program administrators, librarians were guaranteed two class sessions per course. This additional class time allows librarians

to interact with students at multiple points in the research process and incorporate innovative and active learning pedagogies, in which students become actively engaged in assimilating the material instead of passively absorbing it. This two-session model is intended to have a common design so that librarians will be better able to determine the effect of their instruction on student success.

Conducting a multiyear assessment allows librarians to see trends in the data, rather than a snapshot in time of one cohort of students.⁴⁰ This assessment project was repeated in fall 2014 and will be run repeatedly in the future. By continuing to collect and analyze data over multiple years, the librarians are able to “set realistic goals for assessment, based on actual student performance.”⁴¹ This takes away the guesswork when creating benchmarks in programmatic assessments and allows librarians to accurately measure if library instruction is meeting its goals. The “assessment spiral” has been offered as an alternative to the assessment cycle to indicate that assessments should be ongoing but also need to be progressive and improve the quality of student learning instead of circling in a stagnant loop.⁴²

An important part of any assessment project—but especially crucial in an ongoing, multiyear assessment—is reflection.⁴³ Part of that reflection is considering what worked and what needs improvement in the assessment itself.⁴⁴ Multiyear assessments allow the researchers to modify and improve on the assessment approaches, making small changes rather than large overhauls to balance necessary changes with consistency.⁴⁵ After considering the lessons learned from the first round of assessment, the pre- and posttests have been modified to better measure students’ information literacy skill levels, and changes have been suggested to increase response rates. In future rounds, the rubric will be reexamined in an effort to increase ease of use by new raters as well as further improve consistency among them.

Conclusion

This project has brought an increased visibility to information literacy on the campus and positioned the library as a resource for assessment. Participation in this ongoing project has afforded librarians the opportunity to present their findings in a number of venues at the university, both formally and informally. The programmatic nature of the assessment has also been used as a model to develop other measures of student achievement

The librarians have made a number of changes to their instruction and the curriculum as a result of this first round of assessment.



on campus. As the committee members repeat assessments and refine their techniques, the authors expect to move into the future equipped with the skills to document the

Though trying, the experience of designing and implementing a broad baseline assessment from the ground up was also educational and rewarding.

continued impact of their services in a powerful way that, though labor intensive to set up for novices, will yield repeated dividends and help build further expertise. Though trying, the experience of designing and implementing a broad baseline assessment from the ground up was also educational and rewarding. While not every institution will have the same factors pushing and facilitating assessment development, the authors believe that all groups can

successfully implement assessment at some level and hope that the description of and lessons learned by the team at University of Baltimore will make doing so more feasible.

Stephen "Mike" Kiel is a reference and instruction librarian at the Langsdale Library of the University of Baltimore (UB) in Maryland; he may be reached by e-mail at: skiel@ubalt.edu.

Natalie Burclaff is also a reference and instruction librarian at the Langsdale Library of UB; her e-mail address is: nburclaff@ubalt.edu.

Catherine Johnson is head of reference at the United States Naval Academy in Annapolis, Maryland. At the time of writing, she was head of information literacy initiatives at the Langsdale Library of UB. She may be reached by e-mail at: cjohnson@usna.edu.

Notes

1. Christina Leimer, "Organizing for Evidence-Based Decision Making and Improvement," *Change: The Magazine of Higher Learning* 44, 4 (2012): 45.
2. Megan Oakleaf, *The Value of Academic Libraries: A Comprehensive Research Review and Report* (Chicago: Association of College and Research Libraries, 2010): 12–18.
3. Mirna E. Turcios, Naresh Kumar Agarwal, and Linda Watkins, "How Much of Library and Information Science Literature Qualifies as Research?" *Journal of Academic Librarianship* 40, 5 (2014): 478, doi:10.1016/j.acalib.2014.06.003.
4. Ibid.
5. Gila Neta, Russell E. Glasgow, Christopher R. Carpenter, Jeremy M. Grimshaw, Borsika A. Rabin, Maria E. Fernandez, and Ross C. Brownson, "A Framework for Enhancing the Value of Research for Dissemination and Implementation," *American Journal of Public Health* 105, 1 (2015): 50–51, doi:10.2105/AJPH.2014.302206.
6. Deborah Hicks, *Technology and Professional Identity of Librarians: The Making of the Cybrarian* (Hershey, PA: Information Science Reference, 2014), 94.
7. Consuella A. Askew and Eileen Theodore-Shusta, "How Do Librarians Learn Assessment?" *Library Leadership & Management* 28, 1 (2014): 6.
8. M. Lee Upcraft and John H. Schuh, "Assessment vs. Research: Why We Should Care about the Difference," *About Campus* 7, 1 (2002): 17.
9. Megan Oakleaf and Neal Kaske, "Guiding Questions for Assessing Information Literacy in Higher Education," *portal: Libraries and the Academy* 9, 2 (2009): 279.



10. Richael Barger-Anderson, Joseph W. Domaracki, Nedra Kearney-Vakulick, and Richard M. Kubina, “Multiple Baseline Designs: The Use of a Single-Case Experimental Design in Literacy Research,” *Reading Improvement* 41, 4 (2004): 218.
11. Turcios, Agarwal, and Watkins, “How Much of Library and Information Science Literature Qualifies as Research?” 478.
12. Bart L. Weathington, Christopher J. L. Cunningham, and David J. Pittenger, *Research Methods for the Behavioral and Social Sciences* (Hoboken, NJ: Wiley, 2010), 480.
13. Patricia Davitt Maughan, “Assessing Information Literacy among Undergraduates: A Discussion of the Literature and the University of California-Berkeley Assessment Experience,” *College & Research Libraries* 62, 1 (2001): 74.
14. Oakleaf, *The Value of Academic Libraries*, 8
15. Oakleaf and Kaske, “Guiding Questions for Assessing Information Literacy in Higher Education,” 277.
16. Larry H. Ludlow, Joseph Pedulla, Emile Mitescu Reagan, Sarah Enterline, Mac Cannady, and Stephanie Chappe, “Design and Implementation Issues in Longitudinal Research,” *Education Policy Analysis Archives* 19, 11 (2011): 6.
17. Oakleaf and Kaske, “Guiding Questions for Assessing Information Literacy in Higher Education,” 278–81.
18. Ludlow, Pedulla, Reagan, Enterline, Cannady, and Chappe, “Design and Implementation Issues in Longitudinal Research,” 7.
19. Brandy Whitlock and Julie Nanavati, “A Systematic Approach to Performative and Authentic Assessment,” *Reference Services Review* 41, 1 (2013): 41–43, doi:10.1108/00907321311300866.
20. Judith T. M. Gulikers, Theo J. Bastiaens, and Paul A. Kirschner, “A Five-Dimensional Framework for Authentic Assessment,” *Educational Technology Research and Development* 52, 3 (2004): 69.
21. Whitlock and Nanavati, “A Systemic Approach to Performative and Authentic Assessment,” 36.
22. Terrel L. Rhodes and Ashley Finley, *Using the VALUE Rubrics for Improvement of Learning and Authentic Assessment* (Washington, DC: Association of American Colleges & Universities, 2013): 6
23. Dannelle D. Stevens and Antonia Levi, *Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback, and Promote Student Learning* (Sterling, VA: Stylus, 2013), 39–45.
24. Robin Tierney and Marielle Simon, “What’s Still Wrong with Rubrics: Focusing on the Consistency of Performance Criteria across Scale Levels,” *Practical Assessment, Research & Evaluation* 9, 2 (2004): 1–10.
25. Mark S. Litwin, *How to Assess and Interpret Survey Psychometrics* (Thousand Oaks, CA: Sage Publications, 2003), 57–58.
26. Neta, Glasgow, Carpenter, Grimshaw, Rabin, Fernandez, and Brownson, “A Framework for Enhancing the Value of Research for Dissemination and Implementation,” 51.
27. John M. Malouff, Ashley J. Emmerton, and Nicola S. Schutte, “The Risk of a Halo Bias as a Reason to Keep Students Anonymous during Grading,” *Teaching of Psychology* 40, 3 (2013): 236.
28. Claire Holmes and Megan Oakleaf, “The Official (and Unofficial) Rules for Norming Rubrics Successfully,” *Journal of Academic Librarianship* 39, 6 (2013), 599.
29. Anders Jonsson and Gunilla Svingby, “The Use of Scoring Rubrics: Reliability, Validity, and Educational Consequences,” *Educational Research Review* 2, 2 (2007): 134.
30. Holmes and Oakleaf, “The Official (and Unofficial) Rules for Norming Rubrics Successfully,” 599–602.
31. William K. Tschepikow, “Why Don’t Our Students Respond? Understanding Declining Participation in Survey Research among College Students,” *Journal of Student Affairs Research and Practice* 49, 4 (2012): 457.



32. Ibid.
33. Chenicheri Sid Nair, "The Effect of Email Notification on Web-Based Questionnaire Responses," *Journal for Quality & Participation* 35, 4 (2013): 21.
34. Barbara Gross Davis, *Tools for Teaching* (San Francisco: Jossey-Bass, 2009), 395–97.
35. Litwin, *How to Assess and Interpret Survey Psychometrics*, 57–67.
36. Ludlow, Pedulla, Reagan, Enterline, Cannady, and Chappe, "Design and Implementation Issues in Longitudinal Research," 7.
37. Bonnie J. M. Swoger, "Closing the Assessment Loop Using Pre- and Post-Assessment," *Reference Services Review* 39, 2 (May 2011): 245; Megan Oakleaf, "The Information Literacy Instruction Assessment Cycle: A Guide for Increasing Student Learning and Improving Librarian Instructional Skills," *Journal of Documentation* 65, 4 (2009): 545.
38. Oakleaf, "The Information Literacy Instruction Assessment Cycle," 545.
39. Linda Suskie, *Assessing Student Learning: A Common Sense Guide* (San Francisco: Jossey-Bass, 2009), 50.
40. Margaret Fain, "Assessing Information Literacy Skills Development in First Year Students: A Multi-Year Study," *Journal of Academic Librarianship* 37, 2 (March 2011): 118.
41. Ibid.
42. Catherine M. Wehlburg, "Closing the Feedback Loop Is Not Enough: The Assessment Spiral," *Assessment Update* 19, 2 (March 2007): 1; Suskie, *Assessing Student Learning*, 50.
43. Swoger, "Closing the Assessment Loop Using Pre- and Post-Assessment," 251–52.
44. Suskie, *Assessing Student Learning*, 319.
45. Ibid., 321.