Evaluating Library Instruction: Measures for Assessing Educational Quality and Impact

Katherine Schilling and Rachel Applegate

Introduction
This project analyzed data from information literacy research that sought to determine the best sources of information about the effectiveness of information-retrieval instruction and library educational programming. Data from surveys, written tests, and a practical literature searching exercise were correlated to examine the extent to which students’ perceptions about their skills matched their demonstrated skills.

Project data were gathered through a program that examined the impact of information-skills training on medical students’ MEDLINE searching skills, and on their perceptions about their university’s health sciences library. Information-skills training and information literacy are integrated throughout this School of Medicine’s curriculum, with librarians routinely participating in courses as instructors and facilitators. With medical education focused on web technologies for delivering course content to geographically dispersed students, an IL web tutorial was studied to assist clinical teams in decisions about implementing web-based instruction across the curriculum for improved problem-solving, critical thinking, and clinical skills. At the same time, academic libraries had implemented stand-alone, web-based information literacy tutorials, literature was available, but few evaluated web learning tools for delivering curriculum-integrated information training and IL education.

Quality educational program evaluation includes both quantitative and qualitative measures. Many commonly used tools (surveys) focus more exclusively on students’ attitudes about their own skills, expertise, and the library. These are important areas of investigation: user satisfaction and self-confidence are significant factors in understanding students’ attitudes about information usage and information literacy. However, librarians are also challenged to identify evaluation tools that will effectively measure what it is that we want to know about our programming. In addition, we must distinguish between and appropriately apply non-measures (attitudes), with indirect measures (student self-assessment of skills), with direct measures (tests).

Librarians feel strongly about positive professional relationships among librarians and clients, with educational programming recognized as an effective way to initiate and maintain positive contact. Students who meet with librarians or receive library training report less library anxiety, and are more positive about using

Katherine Schilling is Assistant Professor, Indiana University School of Library & Information, email: katschil@iupui.edu. Rachel Applegate is Assistant Professor, Indiana University Purdue University Indianapolis, email: rapplega@iupui.edu.
the library and its resources. Ren found that self-efficacy in electronic searching increased after students participated in course-based library program. Students' attitudes and emotional experiences were positively impacted as a result of the training.

User satisfaction studies report on the multiple factors that are responsible for students' perceptions about library quality: range of materials, helpfulness and availability of library staff, accessibility and ease-of-use of electronic resources, and required skill levels for using resources. With so many variables impacting on students' library use, skills, and information literacy, it becomes difficult to ascertain which training evaluation methods are most effective in any given circumstance. Librarians are continually challenged to identify and apply the most appropriate evaluation methods in-context.

This research investigated the efficacy of several standard evaluation measures including attitudinal surveys, self-assessment of skills, written tests, and a graded literature searching exercise. In order to determine to what extent students' perceptions correlated with their demonstrated skills, self-reported information skills were measured against students' grades on the MEDLINE exercise.

**Literature Review**

The current status of literature on evaluating information-skills education remains limited. The authors identified 33 peer-reviewed articles about university information-skills education published in 2006. Just over half of these articles (18) did not involve training evaluation. Another 14 articles included at least one form of training evaluation, with self-reported attitude surveys being the most frequently used method for evaluating training (reported in 9 articles). Librarians reported using citation pattern analysis, narrative reflection, and focus groups to evaluate library-based training. 'Authentic evaluation' activities such as knowledge tests or graded, course-related activities were also implemented in five instances. Among these publications was a study that reported on 'process implementation evaluation,' the process of surveying faculty feedback on training 'success' versus how well students learned the required materials. Fafeita surveyed Australian vocational libraries, finding that teaching librarians most commonly used these assessment methods.

55% (of librarians) used collaborative learning exercises, 40% used short answer quizzes, 34% used peer and self evaluation, 27% used quizzes, 20% used multiple choice questions, 3% used essays, 2% used diary or journal of search process, and 0% used portfolios.

Table 1 illustrates these findings. Correlating results from multiple types of evaluations can provide useful information to help us understand what it is that evaluations tools—surveys, written tests, practical exercises, etc.—measure best, or do not actually measure at all. Library educational programming articles have reported implementing various forms of training evaluation, but in few cases were these evaluation results cross-correlated to analyze efficacy. Bronshteyn and Baladad assessed learning with individual narrative reflections and anonymous surveys; but these could not be correlated because it was impossible to track anonymous surveys. As these findings illustrate, libraries did not typically implement 'consequential' assessment measures directly related to students' course grades. In fact, the most frequently used evaluation method was to ask (survey) students to rate their own competence or knowledge.

Academic assessment protocols strongly urge the incorporation of 'direct' testing measures, such as objective tests or expert reviews of performances, artifacts or portfolios. Self-reported measures are considered 'indirect': informative, but not specific nor as useful as direct measures. Many process measures, including those in which faculty are surveyed about what should be incorporated into an instructional session, are considered 'non-measures' for the purpose of evaluating of student learning.

Librarians may accept that students are excellent self-judges, but data analysis has often demonstrated otherwise, revealing three important findings: 1) students believed themselves to be more information literate than actual test scores would indicate, 2) students were not aware of important quality-filtered resources, 3) students did not know what information resources to use in support of their coursework. Previous research demonstrated that students assume they possess a higher level of information-literacy skills than they actually have. Plutchak identified the 'satisfied but inept' phenomenon, showing that most end-users were able to identify only a handful of available and applicable resources, went about information retrieval in a haphazard and inefficient manner, and wasted time. Because these...
Everyone found something that seemed relevant, however, they were satisfied with their work.\textsuperscript{36-38}

Nevertheless, there is evidence for a relation between indirect and direct measures. A 2006 report concluded that self-reports have demonstrated a ‘very good relationship’ to objective measures; and therefore, are considered to be valid.

Although results vary depending on the traits and instruments examined, these studies report correlations of .50 to .70, on average, between self-reports and such objective criterion measures as the ACT Comprehensive Test, the College Basic Academic Subjects Examination, and the Graduate Record Examination.\textsuperscript{39}

When direct measures are expensive, cumbersome to implement, or difficult to analyze, indirect measures may be a viable alternative. However, researchers and information literacy instructors must avoid confusing ‘material’ satisfaction (whether the learner demonstrated effective skills) versus ‘emotional’ satisfaction (how the learner felt about his or her skills.) One must also consider the issue of validity of measures, or the process of assessing what should be measured, and which of the many available methods best suits each need given the logistical circumstances.\textsuperscript{40-41}

### Methodology

Using data from surveys, written tests, and literature searching exercise, the authors examined the degree to which students’ perceptions about their skills actually matched their demonstrated skills. Students’ attitudes about the library and librarians were also correlated with a graded exercise to determine whether a relationship existed between attitudes and grades. Data analyses revealed discrepancies between students’ opinions, self-assessments and graded activities, providing insight the most reliable sources of information about the effectiveness of information literacy instruction.

### Table 1: 2006 Peer-Reviewed Journal Articles Evaluating Library Training

<table>
<thead>
<tr>
<th>Citations</th>
<th>Evaluation Method</th>
<th>Assessment Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rui/portal</td>
<td>Citation patterns</td>
<td>Artifacts / expert assessment</td>
</tr>
<tr>
<td>Rui/portal</td>
<td>Course grade</td>
<td>Artifacts / expert assessment</td>
</tr>
<tr>
<td>Rui/portal</td>
<td>Paper grade</td>
<td>Artifacts / expert assessment</td>
</tr>
<tr>
<td>Schaik et al./EIL</td>
<td>Tasks/speed and accuracy</td>
<td>Artifacts / expert assessment</td>
</tr>
<tr>
<td>Lindsay et al./C&amp;RL</td>
<td>Tasks/think-aloud</td>
<td>Artifacts / expert assessment</td>
</tr>
<tr>
<td>Novotny &amp; Cahoy/portal</td>
<td>Tasks / Think-aloud</td>
<td>Artifacts / expert assessment</td>
</tr>
<tr>
<td>Schaik et al./EIL</td>
<td>Knowledge test</td>
<td>Quiz – not part of grade</td>
</tr>
<tr>
<td>Crawford/JLIS</td>
<td>Survey (name databases)</td>
<td>Quiz – not part of grade</td>
</tr>
<tr>
<td>Jackson/C&amp;RL</td>
<td>Quiz</td>
<td>Quiz – part of grade</td>
</tr>
<tr>
<td>Bronshteyn &amp; Baladad/JAL</td>
<td>Narrative reflections</td>
<td>Self-assessment</td>
</tr>
<tr>
<td>Bronshteyn &amp; Baladad/JAL</td>
<td>Survey</td>
<td>Self-assessment</td>
</tr>
<tr>
<td>Wong et al./JAL</td>
<td>Survey</td>
<td>Self-assessment</td>
</tr>
<tr>
<td>Sutton-Knight/JAL</td>
<td>Survey</td>
<td>Self-assessment</td>
</tr>
<tr>
<td>Flatley-Jefferson/LP&amp;P</td>
<td>Survey</td>
<td>Self-assessment</td>
</tr>
<tr>
<td>Brothers-Richardson/S-L</td>
<td>Survey</td>
<td>Self-assessment</td>
</tr>
<tr>
<td>Crawford/JLIS</td>
<td>Survey</td>
<td>Self-assessment</td>
</tr>
<tr>
<td>Graves &amp; Desai/RSR</td>
<td>Survey</td>
<td>Self-assessment</td>
</tr>
<tr>
<td>Desai &amp; Graves/RSR</td>
<td>Survey</td>
<td>Self-assessment</td>
</tr>
<tr>
<td>Schaik et al./EL</td>
<td>Survey (confidence)</td>
<td>Self-assessment</td>
</tr>
</tbody>
</table>
Data Sources
The data sources for this project came from a randomized, double-blinded study that analyzed the impact of web-learning technologies on an established information literacy curriculum for medical students at a large urban university. The information-seeking skills and patterns of first-year medical students enrolled in a required, problem-based learning (PBL) course were tracked. All first-year students participated in a PBL course (N=128) that used small-group study of patient cases to promote diagnostic skills and scientific decision making. Academic health sciences librarians routinely participated in curriculum development and course activities, teaching the information-skills sections of the PBL course, and acting as small-group information liaisons. In addition to being essential to this course, information-seeking, analysis and evaluation are core components of PBL in medical education.

Students were randomly assigned to experimental groups. The control group (n=63) participated in traditional, instructor-lead, information-seeking training; and the intervention group (n=65) participated in identical instruction via a web tutorial. The instructor-lead (control) and e-learning (intervention) sessions were anchored in ACRL information literacy competency standards, and on several information-retrieval models.

Immediately prior to information-retrieval training, students completed a pre-training survey with skills self-assessment, attitudinal scales, and a written test section. Control and intervention groups then moved to classrooms for 90-minute training. During the next week, students completed a homework assignment that required them to perform a MEDLINE search on a course-related topic. Pre- and post-surveys, MEDLINE searches, and other research data were tracked using individual account login names. Searches were electronically captured, blinded, and independently evaluated by three expert searchers who routinely participated in the PBL course.

Two weeks later, a post-training survey was administered during class. The post-survey mirrored the pre-survey, also including a training evaluation section and questions about which information resources students expected to use for the PBL course during the upcoming semester.

The pre- and post-training survey instruments included three parallel sections consisting of a questionnaire, self-assessment, and test. The surveys asked participants to report their perceptions regarding how the library training impacted on their comfort levels, attitudes, and projected library use. The post-survey asked a variety of questions in which respondents provided feedback about the training experience itself. Survey data were also correlated with students’ literature searching scores to identify potential relationships between attitudes and performance. A three-month follow-up survey gathered additional feedback on students’ attitudes and presented a picture of their overall information usage patterns over the course of the semester.

Results
This project focused on deepening our understanding of what evaluation factors are most efficacious and garner the most useful and reliable data about students’ use of information and library resources. In this project, multiple training evaluation (surveys, self-assessments, written tests, and an expertly graded assignment) were cross-correlated with performance scores. These analyses sought to identify ‘gold standard’ variables for evaluating information-literacy training and educational programming.

Data analysis focused on three groups of variables (expert evaluations, students’ self-assessments, students’ attitudes) with data gathered from pre- and post-training survey questions:

(E) Expert Evaluations
1. Test scores (in the form of a post-tutorial survey; ungraded)
2. MEDLINE exercise independently evaluated by expert librarians

(J) Student Self-Assessment (students as self-judges):
3. Student Assessment of Workshop
   “How effective was this workshop in improving…skills?” (p11)
4. Student Satisfaction with Search Results
   “How satisfied are you with the results of” the search….? (p15)
5. Student Self-Assessment of skills
   “How skilled do you consider yourself to be…?” (p33)

(A) Student attitudes
6. Student Exercise Attitude-Satisfaction
   “The Medline search exercise left me satisfied” (p14h)
7. Student Exercise Attitude-Confidence
   “The Medline search exercise left me confident” (p14e)
8. Student Exercise Attitude-Frustration
   “The Medline search exercise left me frustrated” (p14e)

E=Expert evaluations : J=Students as judges (EJ)
E=Expert evaluations : A=Students report attitudes (EA)
J=Students judge their work : A=Students report attitudes (JA)

Skills Self-Assessments
Only learners themselves can be accurate reporters of their emotions, satisfaction, confidence, or frustration. An important question is whether they can be accurate judges of other issues such as the effectiveness of the workshop, their own skills, or material satisfaction (whether the search was correct and effective). Data analyses revealed that only one of the three major EJ correlations was statistically significant, and that the EA and JA correlations were not statistically significant. Correlations were tested against a p < 0.05 criterion.

Expert Evaluation vs. Student Judgment: Statistically significant correlation.
Search scores and student assessment of workshop
“How effective was this workshop in improving…skills” (p11)
EJ +0.25 statistically significant (s.s.)
The correlation above revealed that students who believed that their skills had improved because of the information-seeking workshop, actually did demonstrate higher skills than those students who did not believe that their skills had improved as a result of training. Students were to be believed, then, when they reported that participating in training had a real and positive impact on their skills development.

Expert Evaluation vs. Student Judgment: Not statistically significant
Search scores and student assessment of skills
“How skilled do you consider yourself to be” (p33)
EJ -0.07 (n.s.)

The absence of a correlation above showed that students who rated themselves as being more highly skilled did not actually demonstrate a higher level of skill on a graded exercise. When compared with other data, findings revealed that students accurately predicted whether they had learned something, but did not accurately predict their own actual skill levels. This reiterates the concept of the ‘satisfied but inept’ user, reminding us that people often believe themselves to be more information literate than they actually are.48 Additional data analyses showed no relationship between students’ searching exercise grades and their attitudes about the library, librarians, and information resources, demonstrating a disconnect between grades and satisfaction. Students who performed poorly on the exercise did not necessarily feel more negatively about the library resources and librarians. On the other hand, this also meant that students who performed well on the exercise did not necessarily feel more positively about the library.

Search scores and student satisfaction with search
“How satisfied are you with the results of” the search (p15)
EJ +0.06 (n.s.)

Demonstrated Skills and Self-Reported Attitudes
Expert Evaluations vs. Student Attitudes: Not statistically significant
Search scores and student exercise attitude-confidence
“The Medline search exercise left me confident” (p14a)
EA +0.12 (n.s.)

Search scores and student exercise attitude-frustration
“The Medline search exercise left me frustrated” (p14e)
EA -0.09 (n.s.)
Search scores and student exercise attitude-satisfaction

“The Medline search exercise left me satisfied” (p14h)

EA +0.13 (n.s.)

Data analyses found no correlation between expert evaluations and students’ attitudes, above. These results showed that neither 1) self-reported confidence nor 2) satisfaction with the results of their MEDLINE homework accurately predicted students’ grades. Students who were highly confident or satisfied with their assignments did not achieve higher searching scores than did those students who were less confident or less satisfied.

Interestingly, there was a divergence between confidence and satisfaction: “This exercise left me confident” did not significantly correlate with “This exercise left me satisfied.” As one would expect, however, the more frustrated a student reported being, the less satisfied and the less confident he/she also reported feeling (r=-0.19 and -0.33 respectively). Perhaps confidence does not beget satisfaction, but a lack of confidence does beget dissatisfaction.

Students as Judges of Their Own Learning and Attitudes

Statistically significant correlations were identified between students’ judgments and satisfaction attitudes. Students who were more confident about their skills were also happier with their work; students who were less confident about their skills were also less happy with their work.

Student Judgments vs. Student Attitudes: Statistically Significant

JJ +0.22 (s.s.)

Student assessment of skills and student satisfaction with search results

“How skilled do you consider yourself to be” (p33)

“How satisfied are you with the results of” the search (p15)

JJ +0.22 (s.s.)

Student Judgments and Student Attitudes: Statistically Significant

(in order of effect size)

Students satisfaction with search results and student exercise attitude-frustration

“How satisfied are you with the results of?” the search (p15)

“The Medline search exercise left me frustrated” (p14e)

JA -0.44 (negative s.s.)

Student assessment of workshop and student exercise attitude-satisfaction

“How effective was this workshop in improving…skills” (p11)

“The Medline search exercise left me satisfied” (p14e)

JA +0.41 (s.s.)

Student assessment of workshop and student satisfaction with search results

“How effective was this workshop in improving…skills” (p11)

“How satisfied are you with the results of” the search (p15)

JA +0.23 (s.s.)

Student satisfaction with search results and student exercise attitude-satisfaction

“How satisfied are you with the results of” the search (p15)

“The Medline search exercise left me satisfied” (p14h)

JA +0.19 (s.s.)

Student satisfaction with search results and student exercise attitude-confidence

“How satisfied are you with the results of” the search (p15)

“The Medline search exercise left me confident” (p14e)

JA +0.23 (s.s.)

Written Tests vs. Demonstrated Searching Skills

Pre- and post-survey written tests evaluated students’ abilities to apply, on paper, skills such as correctly applying Boolean operators; distinguishing between searchable fields; identifying various publication types; distinguishing between various citation types and formats; and other questions designed to assess their familiarity with information literacy concepts. Data analyses showed no correlation between written and applied (MEDLINE exercise) abilities. These findings indicated a disconnect between written versus applied skills, showing that a knowledge of information literacy concepts did not translate into practical skills in maneuvering through a sophisticated bibliographic database.

Library and Information-Seeking Skills

Post- survey data were compared with end-of-the-semester library usage statistics and follow-up survey data. Results showed that coming into the course, students believed that they were information literate. Analyses
also found that despite their confidence, students were not aware of core resources, and did not know which information resources were most appropriate or useful for their PBL course. These findings present an important issue for educational intervention: librarians must find ways to overcome learners’ beliefs that their skills are perfectly adequate and convince learners that information-skills training is necessary and useful.

Discussion

Educators and academic librarians know that information literacy skills are vital to academic success. The transference of these skills via library programming improves students’ information skills and self-sufficiency. The results of data analyses to identify relationships between students’ attitudes and performance on a scored exercise resulted in very few statistically significant correlations. This research found that for the most part, students were not particularly good at assessing their own information skills. Students did know when they learned something, and were good reporters of their own feelings and attitudes. These attitudes did not correlate with students’ demonstrated skills, however. Research results also showed that comfort indicators did not forecast or correspond to demonstrated skill.

Students demonstrated or practical information skills provided perhaps the most accurate and objective information about the efficacy of information training. Although attitudes did not measure skills or correlate with skills, attitudinal tools do remain important in the larger context, however. Librarians want students to feel comfortable and confident in the library, but also learn to identify their own limitations, knowing when to seek out assistance from a librarian. Users’ attitudes about library resources create a delicate balance: when users need assistance AND have the right attitude, they actually will seek assistance. When users need assistance AND do not have the right attitude, they will not seek assistance. Users who are unjustifiably over-confident may demonstrate an unrealistic ‘I can do it all by myself’ attitude about library research. In a related study, focus group students commented that they felt like they were “cheating” if they asked librarians for help with information research.49

This and related research illustrate the potential for more accurate and realistic applications for evaluating library training. Self-reported surveys, for example, were an excellent option for gleaning a better understanding of students’ feelings, but did not reveal anything about students’ skill levels or information usage habits. Written, theoretical tests did not accurately reveal students’ skills with the same level of clarity and accuracy as did the practical exercise. Students who said that they HAD learned something, DID actually learn something. However, students may have assumed that their skills were higher than they actually were. Despite students’ confidence in their own skills, they were not fully prepared for the information literacy requirements of their course. Findings are summarized in Table 3.

Conclusion

Extensive experience and research have demonstrated that early and sequential educational intervention impacts positively on students’ abilities to effectively find and use information resources, and on their overall information literacy development. As libraries continue to invest significant time and attention to educational programs, effective evaluation is increasingly important. Verifying the effectiveness of interventions requires academic librarians to decide exactly what it is that we want to measure, and apply evaluation tools to determine whether the intervention resulted in measurable changes. This project found that different evaluation tools were not equally effective in measuring learning variables such as students’ skills, attitudes, or library and information usage. These results may provide valuable information about the reliability of commonly-used evaluation tools in library educational programming.

Notes

Table 3: Summary Findings

<table>
<thead>
<tr>
<th>Evaluation Tool</th>
<th>Measured</th>
<th>Did NOT Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practical exercises</td>
<td>• Practical, applied skills</td>
<td>• Library usage / information usage</td>
</tr>
<tr>
<td>(graded homework; evaluated, course-related assignments)</td>
<td></td>
<td>• Attitudes, feelings, comfort, beliefs, perceptions</td>
</tr>
<tr>
<td>Written test</td>
<td>• Overall familiarity with concepts, vocabulary, theoretical applications</td>
<td></td>
</tr>
<tr>
<td>(ungraded, on-paper, written skills test)</td>
<td></td>
<td>• Practical, applied skills</td>
</tr>
<tr>
<td>Self-reported self-assessment scales</td>
<td>• Confidence</td>
<td>• Overall familiarity with concepts, vocabulary, theoretical applications</td>
</tr>
<tr>
<td>(‘I am good at / I am not good at’ ‘I am skilled / I am not skilled’ ‘I am experienced / I am not experienced’)</td>
<td></td>
<td>• Library usage / information usage</td>
</tr>
<tr>
<td>Attitudinal surveys</td>
<td>• Attitudes, feelings, comfort, beliefs, perceptions</td>
<td>• Overall familiarity with concepts, vocabulary, theoretical applications</td>
</tr>
<tr>
<td>(opinions, feelings, preferences, likes/dislikes)</td>
<td></td>
<td>• Library usage / information usage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Overall familiarity with concepts, vocabulary, theoretical applications</td>
</tr>
</tbody>
</table>


33. K. Bronshteyn and R. Baladad. “Librarians as Writing Instructors.”

34. C. A. Palomba and T. W. Banta. *Assessment Essentials*.


