

9-c: Digital Library Evaluation, User Studies

Draft, Barbara Wildemuth, 11/16/2007

1. Module name: Digital Library Evaluation, User Studies

2. Scope

While a number of kinds of evaluation/research studies may be conducted during the design and development of a digital library (e.g., usability testing), this module is concerned with methods for evaluating the outcomes, impacts, or benefits of a digital library, including cost/benefit analyses. It also includes methods that are useful for general user studies (i.e., studies that intend to more fully understand people's interactions with digital libraries). While some methods covered here are useful for usability testing, usability inspections and usability testing are explicitly covered in module 6-d, Interaction Design, Information Summarization and Visualization, and Usability Assessment.

3. Learning objectives

Students will be able to:

- Understand the importance of DL evaluation;
- List and describe the strengths and weaknesses of multiple approaches to evaluation; and
- Apply an appropriate evaluation method to a particular DL.

4. 5S characteristics of the module

Streams: N/A

Structures: N/A

Spaces: N/A

Scenarios: Scenarios may form the basis of an evaluation plan, by describing particular situations of use that must be supported effectively by the DL.

Societies: The concept of societies may be useful in planning an evaluation because it will support the evaluator in more systematically consider the potential stakeholders of the DL.

5. Level of effort required (in-class and out-of-class time required for students)

- In-class time: 2-2½ hours
- Out-of-class time: 1½ hours for assigned reading
- Learning activities (optional): See notes on timing with each activity or assignment.

1 **6. Relationships with other modules (flow between modules)**

2 It is expected that this module will follow other modules on digital libraries, and will
3 be in the final portions of the module sequence.
4

5 **7. Prerequisite knowledge required (completion optional)**

6 Students will *not* be expected to have had prior training in social science research
7 methods.
8

9 **8. Introductory remedial instruction: None**
10

11 **9. Body of knowledge**

12 *Evaluation and user studies:*

13
14 Definition of evaluation: “An appraisal of the performance or functioning of a
15 system, or part thereof, in relation to some objective(s)” Saracevic, 2000,
16 p.359
17

18 Evaluation incorporates the making of value judgments about whether
19 performance is adequate
20 Its purpose is to inform decision making (Reeves et al., 2003)
21

22 Evaluation is critical to any project
23 NSF recommended that at least 10% of the project budget be devoted to
24 evaluation in their early DL initiatives
25

26 User studies may be more general, in terms of the types of questions asked
27 They do not necessarily incorporate the making of value judgments about
28 performance quality
29 User studies may be more specific, in that they involve users
30 Evaluations may be conducted on the DL collection or other aspects of the
31 DL without involving users
32

33 This module will focus particularly on evaluations that involve users
34

35 *The object of the evaluation or user study: digital libraries and their*
36 *processes/functions*

37
38 A particular aspect of a digital library
39 Individual digital libraries
40 Multiple digital libraries
41

- 1 DL processes that may be evaluated (based on Saracevic, 2005), and criteria for
- 2 evaluation
- 3 Information representations, metadata and surrogates used in the DL
- 4 Task appropriateness
- 5 Usability
- 6 User satisfaction
- 7 Particular tools available in the DL
- 8 Index, search, and output features
- 9 Navigation, browsing
- 10 Failures in functionality or usability
- 11 User satisfaction
- 12 Particular services offered by the DL
- 13 Collection quality
- 14 Retrieval performance (recall, precision)
- 15 Reliability
- 16 Human-intermediated services (e.g., reference services)
- 17 User satisfaction with individual services or with collection of
- 18 services
- 19 User behaviors when interacting with the DL (may or may not be
- 20 evaluative)
- 21 Information seeking/searching behaviors
- 22 Use of information retrieved
- 23 Work patterns
- 24
- 25 “All efforts to design, implement, and evaluate digital libraries must be rooted in the
- 26 information needs, characteristics, and contexts of the people who will or may use
- 27 those libraries.” Marchionini, Plaisant, & Komlodi, 2003, p.1
- 28

29 *Questions that may be asked during an evaluation/user study*

- 30
- 31 Frame the study questions based on the decisions that must be made about the
- 32 DL’s functions/processes (Reeves et al., 2003)
- 33 Focus on those questions that are most important for making the decisions
- 34 that are most important
- 35 Focus on impacts of DL functions/services (Marchionini, Plaisant, &
- 36 Komlodi, 2003)
- 37 What types of impacts are there? On whom?
- 38 Who and what influence those impacts?
- 39
- 40 Formative versus summative evaluation
- 41 Formative evaluation focused on decisions about how to modify/change
- 42 the DL’s functions/services
- 43 Summative evaluation focused decisions about the worth or value of the
- 44 DL’s functions/services
- 45
- 46

1 ***Stages/steps in the evaluation/research process***

2
3 Develop an evaluation/research plan (Chowdhury & Chowdhury, 2003; Reeves et
4 al., 2003)

5 Clarify the decisions to be addressed and the questions they generate

6 Identify appropriate evaluation methods, including sampling procedures,
7 data collection procedures and data analysis procedures

8 Carry out the evaluation/research plan

9 Report the results to appropriate stakeholders

10 Primarily the decision makers

11 Also other constituencies of the DL

12
13 Review of an example evaluation, in terms of these steps

14 Possible example evaluations:

15 Bishop, A.P. (1998, December). Measuring access, use, and success in
16 digital libraries. *The Journal of Electronic Publishing*, 4(2). Retrieved
17 February 8, 2006, from www.press.umich.edu/jep/04-02/bishop.html.

18 Marchionini, G. (2000). Evaluating digital libraries: A longitudinal and
19 multifaceted view. *Library Trends*, 49(2), 304-333.

20
21 ***Evaluation design strategies***

22
23 Naturalistic studies

24 For some evaluation studies, it is critical to conduct them in a natural or
25 naturalistic setting/context

26 The constraints of the setting usually imply that fewer experimental
27 controls can be applied to the study design

28 Usually, the evaluator will need to take into account aspects of the setting
29 as part of the data collected for the evaluation study

30
31 Experiments

32 Usually conducted in a lab setting, or a setting in which control over the
33 conditions of the evaluation study can be exerted

34 The researcher attempts to control all the potential effects on the results of
35 the study, other than those effects being intentionally manipulated as
36 the focus of the evaluation

37
38 Some important concepts in designing an experiment

39 Randomization is a key tool for control: random sampling and/or
40 random assignment to treatment and control groups

41 Variables: the researcher will manipulate the independent variables
42 (e.g., whether the DL has a particular feature or not) and will
43 evaluate the outcomes based on the dependent variable

1 The design may be a within-subjects design (where each
2 participant interacts with all the variations of the independent
3 variables, and so comparisons on the dependent variable are
4 made “within” each subject’s performance) or a between-
5 subjects design (where each participant interacts with only one
6 version of the system and comparisons are made “between”
7 groups of subjects)
8

9 **Avoiding the effects of researcher bias**

10
11 It’s easy for a researcher’s biases to influence the design of a study and,
12 thus, its outcomes
13 Identify your biases
14 Ensure that your study design and procedures will allow you to
15 avoid any influence on the study outcomes
16

17 ***Data collection and measurement methods***

18
19 Collecting data from people requires ethical treatment of those people as study
20 participants

21 Each institution will require review of the research proposal by an
22 Institutional Review Board that verifies that study participants are
23 being treated ethically
24

25 **Observation of user behaviors, including transaction logs**

26
27 To see what the user is doing as he or she interacts with the system
28

29 **Observation of work (e.g., via contextual inquiry)**
30 Special kind of interview
31

32 Observe person while performing task to be supported
33 Interrupt with questions about how and why, as needed
34

35 **Think-aloud protocols**

36 During DL use, the participant is asked to verbalize their thought
37 processes

38 Allows you to observe “unobservable” cognitive behaviors
39 Usually videotaped or audiotaped
40

41 **Indirect observation of work**

42 Logging and metering techniques embedded in the software of the
43 current system or intermediate versions
44
45

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

Diaries

- For detailed descriptions of tasks
- How much time they take
- Sequential dependencies between tasks

Allows observation over longer periods of time than contextual inquiry interviews

Interviews and focus groups

Augmenting other data collection methods, or on their own

Uses during DL evaluation

- For identifying problems in the DL design
- For additional features needed in the DL
- For other improvements in the DL which the user can suggest

Individual interviews or group interview (focus groups)

Focus groups require a skilled facilitator

Questionnaires

Surveys: typically one item/question per construct

Measures: intended to measure constructs that are not directly observable and not easily measured with a single item

The more subjective the construct, the more likely that you will need a multiple-item measure for it

Find a measure in the literature, rather than developing your own

Print vs. online administration

Print allows people to annotate (can be good or bad)

Online eliminates the need for a separate data entry step

Study sample: Who should be participants in your evaluation study?

Define the population of interest

Current users or a subset of them

Potential audiences (who are not current users)

Consider sample size

Usually a tradeoff between small sample (cheaper) and generalizability

Intensive versus extensive studies

1 Intensive studies: to thoroughly understand a phenomenon within
2 its context

3 Extensive studies: to understand the extent of a phenomenon
4 within a population

5

6 Develop a sampling plan

7

8 Random sampling

9 Supports statistical inferences to the population

10

11 Identify a population to which you want to generalize your
12 findings

13 Enumerate the population

14 Draw a random sample

15

16 Problem: enumerating the entire population

17 Not necessarily problematic, but often is

18

19 Other methods of sampling

20 Quota sampling, purposive sampling, accidental/convenience
21 sampling

22

23 Strive for representativeness

24 In range, as well as central tendency

25

26 Develop a plan for recruiting the sample you want

27 May need to offer incentives

28

29 ***Analysis and interpretation of data***

30

31 Reporting the results and interpreting the results are two distinct steps

32

33 Interpretation should address the questions, “What do the results mean? How
34 should they be understood?”

35

36 All results must be interpreted in the context of:

37

38 Prior empirical work and relevant theoretical frameworks

39 Situation

40 What is happening in the particular situation in which the study
41 was done?

42

1 Weaknesses in the research

2 Measurements: level of reliability; validity

3 Design: attrition, external events; internal and external threats to
4 validity

5 Analysis method: assumptions violated

6
7 Recommend particular actions, based on the interpretation of the results
8
9

10 **10. Resources**

11 ***Assigned readings for students:***

12 Nicholson, Scott. (2004). A conceptual framework for the holistic measurement and
13 cumulative evaluation of library services. *Journal of Documentation*, 60(2), 164-
14 182.

15 Reeves, Thomas, Apedoe, Xornam, & Hee Woo, Young. (2003). Evaluating digital
16 libraries: A user-friendly guide. University Corporation for Atmospheric
17 Research; National Science Digital Library. Retrieved 3/1/2007

18 from <http://eduimpact.comm.nsdlib.org/evalworkshop/UserGuideOct20.doc>.

19 Chapter 1, Why evaluate? (p.1-6)

20 Chapter 2, Evaluating planning (p.7-21)

21

22 ***Recommended background reading for instructor:***

23 Chowdhury, G.G., & Chowdhury, S. (2003). Digital library evaluation. In
24 *Introduction to Digital Libraries*. London: Facet Publishing, 267-283.

25 Frechtling, J. (2002). *The 2002 User Friendly Handbook for Project Evaluation*.
26 National Science Foundation, Directorate for Education & Human Resources,
27 Division of Research, Evaluation, and Communication.

28 Marchionini, G., Plaisant, C., & Komlodi, A. (2003). The people in digital libraries:
29 Multifaceted approaches to assessing needs and impact. In Bishop, A.P., Van
30 House, N.A., & Battenfield, B.P. (eds.), *Digital Library Use: Social Practice in*
31 *Design and Evaluation*. Cambridge, MA: MIT Press, 119-160.

32 Reeves, Thomas, Apedoe, Xornam, & Hee Woo, Young. (2003). Evaluating digital
33 libraries: A user-friendly guide. University Corporation for Atmospheric
34 Research; National Science Digital Library. Retrieved 3/1/2007

35 from <http://eduimpact.comm.nsdlib.org/evalworkshop/UserGuideOct20.doc>.

36 Saracevic, T. (2000). Digital library evaluation: Toward evolution of concepts.
37 *Library Trends*, 49(2), 350-369.
38

39 ***Additional potential readings include:***

40 Bishop, A.P. (1998, December). Measuring access, use, and success in digital
41 libraries. *The Journal of Electronic Publishing*, 4(2). Retrieved 2/8/2006
42 from www.press.umich.edu/jep/04-02/bishop.html.

- 1 Bishop, A.P., Mehra, B., Bazzell, I., & Smith, C. (2003). Participatory action research
2 and digital libraries: Reframing evaluation. In Bishop, A.P., Van House, N.A., &
3 Battenfield, B.P. (eds.), *Digital Library Use: Social Practice in Design and*
4 *Evaluation*. Cambridge, MA: MIT Press, 161-189.
- 5 Bollen, J. and R. Luce. (2002). Evaluation of digital library impact and user
6 communities by analysis of usage patterns. *D-Lib Magazine*, 8(6) June 2002.
7 Retrieved 3/1/2007 from <http://www.dlib.org/dlib/june02/bollen/06bollen.html>
- 8 Bryan-Kinns, Nick & Blandford, Ann. (2000). A survey of user studies for digital
9 libraries. RIDL Working Paper. Retrieved 3/1/2007
10 from <http://www.cs.mdx.ac.uk/ridl/DLuser.pdf>.
- 11 Choudhury, G.S.; Hobbs, B.; M Lorie, Flores, N.E. (2002, July/August). A
12 framework for evaluating digital library service. *D-Lib Magazine*, 8(7/8).
13 Retrieved 3/1/2007
14 from <http://www.dlib.org/dlib/july02/choudhury/07choudhury.html>.
- 15 Marchionini, G. (2000). Evaluating digital libraries: A longitudinal and multifaceted
16 view. *Library Trends*, 49(2), 304-333.
- 17 Rieger, R., & Gay, G. (1999, June 15). Tools and Techniques in Evaluating Digital
18 Imaging Projects. *RLG DigiNews*.
- 19 Saracevic, Tefko (2005). How were digital libraries evaluated? Presented at Libraries
20 in the Digital Age (LIDA), Dubrovnik and Mljet, Croatia, May 30-June 3.
21 Retrieved 3/1/2007
22 from http://www.scils.rutgers.edu/~tefko/DL_evaluation_LIDA.pdf.
- 23 Thong, J. (2002). Understanding user acceptance of digital libraries: What are the
24 roles of interface characteristics, organizational context, and individual
25 differences? *International Journal of Human-Computer Studies*, 57(3), 215-242.
26

27 **11. Concept Map (created by students)**

28

29 **12. Exercises / Learning activities**

30 **a. Analyze a DL evaluation report**

31 Exercise 13.a, “Analyze a DL evaluation report,” could be adapted to an in-class
32 small-group discussion exercise. If so, the results of the each group’s analysis could
33 be reported orally or could be posted to a class wiki or discussion forum.

34
35 If used as an in-class exercise, assign the groups to read a particular evaluation report
36 before class, to prepare for their in-class discussion and report.

37
38 Time requirements: 2 hours of preparation outside of class; 25-30 minutes for
39 discussion in class; 20-30 minutes for report presentation in class, depending on the
40 number of groups.

41

1 **b. *Develop an evaluation plan***

2 Based on a DL that is familiar to all the students in the class (e.g., flickr, MySpace, a
3 music collection, the university's OPAC or a special collection that is well-known),
4 have students work in small teams (3-4 people each) to develop a draft evaluation
5 plan. They can play the role of an evaluation consulting firm, designing an evaluation
6 study for their client, the DL managers.

7
8 Each plan should include the following:

9
10 The evaluation questions to be addressed

11 Stated briefly, in one sentence (preferably ending with ?)

12 The sample to be included

13 How they will be selected

14 How they will be recruited

15 The methods for data collection

16 The types of data to be collected, and how each pertains to the evaluation
17 question

18 The procedures for collecting the needed data
19

20 Have each team of students present their plan to the class, as if it were an initial
21 presentation to the client (the DL managers).
22

23 Time requirements: It is expected that the students will prepare their report after this
24 module has been presented in class. Students should expect to spend 4-5 hours
25 outside of class, preparing their reports. Each report should be presented in 7-10
26 minutes, during the next class session.
27

28 **c. *Interview a digital librarian about evaluation***

29 Note: This exercise is only possible if there are a number of robust local digital
30 library projects, and the students will have access to their directors/administrators.
31

32 For this exercise, students should work in pairs; each pair will be assigned to
33 investigate a particular digital library. Prior to the interview, each pair should read
34 the available documentation on the digital library on which they're focused. Using the
35 following interview guide, they should interview the director/administrator of the
36 digital library.
37

38 Interview guide:

39 When was the digital library first established?

40 What are the primary goals of the DL?

41 In what ways do you evaluate whether you're achieving those goals?

42 Do you evaluate any other aspects of the DL's operations? If so, how?

43 How and to whom are the evaluation results reported?
44

1 Each pair should write up a brief (1-3 page) summary of their interview findings. In
2 addition, they should be prepared to orally report on the most interesting aspects of
3 those findings at the next class session.

4
5 Time requirements, outside of class: 1-2 hours for preparatory reading; 1 hour for
6 conducting the interview; 2-3 hours for writing up the interview report.

7 Time requirements, in class: 30-40 minutes for the class to discuss the findings from
8 the interviews.
9

10 **13. Evaluation of learning objective achievement**

11 **a. Analyze a digital library evaluation report**

12 Using Saracevic's (2005) meta-analysis of digital library (DL) evaluations as a
13 framework, evaluate an additional DL evaluation report. The report can be selected
14 from the following¹:

15 Byrd, S., et al. (2001). Cost/benefit analysis for digital library projects: The
16 Virginia Historical Inventory project (VHI). *The Bottom Line: Managing*
17 *Library Finances*, 14(2), 65-75.

18 Gambles, A. (2001). The HeadLine personal information environment:
19 Evaluation Phase One. *D-Lib Magazine*,
20 7(3). www.dlib.org/dlib/march01/gambles/03gambles.html.

21 Palmer, D., & Robinson, B. (2001). Agora: The hybrid library from a user's
22 perspective. *Ariadne*, 26. www.ariadne.ac.uk/issue26/case-studies/intro.htm.

23 Zhang, Y., Lee, K., & You, B.-J. (2001). Usage patterns of an electronic
24 theses and dissertations system. *Online Information Review*, 25(6), 370-377.
25 List of possibilities; still need to be viewed.
26

27 Analyze the evaluation report in terms of the following aspects:

- 28 • **“Construct”** for evaluation.
29 What was evaluated? What was actually meant by a “digital library”? What
30 elements (components, parts, processes...) were involved in evaluation?
- 31 • **Context** of evaluation - selection of a goal, framework, viewpoint or level(s)
32 of evaluation.
33 What was the basic approach or perspective? What was the level of evaluation?
34 What was the objective(s)?
- 35 • **Criteria** reflecting performance as related to selected objectives.
36 What parameters of performance were concentrate[d] on? What dimension or
37 characteristic [was] evaluated?
- 38 • **Methodology** for doing evaluation.
39 What measures and measuring instruments were used? What samples? What
40 procedures were used for data collection? For data analysis?
- 41 • **Findings** from evaluation studies
42 Only a single generalization is provided.” (Saracevic, 2005, p.2-3)

¹ Additional reports can be added to this list or reports can be deleted, at the instructor's discretion.

1
2 Prepare a report (2-5 pages, single-spaced) summarizing the findings of your analysis.

3
4 The report should be evaluated in terms of its demonstration that the authors
5 understood the DL evaluation conducted, its coverage of the five aspects of
6 evaluations posed by Saracevic, its identification of strengths and weaknesses in the
7 DL evaluation, and its clarity (organization, grammar, etc.).

8
9 Time requirements: approximately 6-8 hours outside of class.
10

11 **b. Develop an evaluation plan**

12 Class exercise 10.b, “Develop an evaluation plan,” could be adapted as a graded
13 assignment. Each team would be expected to develop their plan over the week after
14 the class’s discussion of evaluation. If class time is available, the final plan can be
15 presented orally; or, if preferred, the evaluation plans could be turned in as an
16 evaluation proposal (2-4 pages).

17
18 The evaluation plans would be evaluated in terms of their completeness (were all the
19 major components of an evaluation study addressed?), their feasibility (could the
20 evaluation study actually be conducted, given reasonable resources?), and their clarity.

21
22 Time requirements: 6-8 hours outside of class, preparing and writing the evaluation
23 plan.
24

25 **14. Glossary**

26
27 Between-subjects design: A research design in which “each research participant receives
28 only one level of the independent variable” (Schmidt, 2000).

29 Dependent variable: “A variable that may, it is believed, be predicted by or caused by one
30 or more other variables called independent variables.” (U.S. Dept. of Justice, n.d.)

31 Evaluation: “An appraisal of the performance or functioning of a system, or part thereof,
32 in relation to some objective(s)” Saracevic, 2000, p.359

33 Formative evaluation: An evaluation that is intended to “strengthen or improve the object
34 being evaluated. Formative evaluations are used to improve [information systems]
35 while they are still under development.” (Trochim, 2001, p.347)

36 Independent variable: “A variable that may, it is believed, predict or cause fluctuation in
37 an dependent variable.” (U.S. Dept. of Justice, n.d.)

38 Research design: “A plan of what data to gather, from whom, how and when to collect
39 the data, and how to analyze the data obtained.” (U.S. Dept. of Justice, n.d.)

40 Sample: “The actual units you select to participate in your study.” (Trochim, 2001, p.351)

41 Stakeholders: “People who have a vested interest in the success of the project or are
42 involved in the implementation of the project.” (California State University,
43 Monterey Bay, n.d.)

1 Summative evaluation: An evaluation that “examine[s] the effects or outcomes of [an
2 information system].” (Trochim, 2001, p.352)
3 Within-subjects design: A research design in which “each research participant provides
4 data for all the levels of the independent variable” (Schmidt, 2000).

5

6 References for glossary:

7 California State University, Monterey Bay, Information Technology. (n.d.) Data
8 warehouse glossary. Retrieved 5/23/2007
9 from <http://it.csUMB.edu/departments/data/glossary.html>.

10 Saracevic, T. (2000). Digital library evaluation: Toward evolution of concepts. *Library
11 Trends*, 49(2), 350-369.

12 Schmidt, S.R. (2000). Research methods for the digitally inclined. Retrieved 11/16/2007,
13 from <http://www.mtsu.edu/~sschmidt/methods/design/design.html>.

14 Trochim, W.M.K. (2001). *The Research Methods Knowledge Base*. Second ed.
15 Cincinnati, OH: Atomic Dog Publishing.

16 U.S. Dept. of Justice, Office of Justice Programs, Bureau of Justice Assistance, Center
17 for Program Evaluation. (n.d.). Glossary. Retrieved 5/23/2007
18 from http://www.ojp.usdoj.gov/BJA/evaluation/glossary/glossary_r.htm.

19

20

21 **15. Additional Useful links**

22

23 None

24

25 **16. Contributors**

26

27 Barbara Wildemuth (UNC, developer), Jeff Pomerantz (UNC, evaluator), Sanghee Oh
28 (UNC, evaluator), Mike Christel (CMU, evaluator), Flora McMartin (Broad-based
29 Knowledge, evaluator), Liz Liddy (Syracuse, evaluator)

30