Evaluating the User experience: What to Ask, How to Measure, and What to Learn from Assessment

Joyce Chapman, State Library of North Carolina
What is assessment?
Considerations for data collection
Harnessing operational data
Techniques for evaluation
  - A/B testing
  - Cost/benefit analysis
What *is* assessment?

Assessment is a continuous and cyclical process by which we evaluate and improve services, products, workflows, and learning.
What is assessment?

Assessment is a continuous and cyclical process by which we evaluate and improve services, products, workflows, and learning.
WHAT IS ASSESSMENT?

Assessment cycle

Plan

Implement

Analyze / report

React / refine
Considerations for data collection
CONSIDERATIONS FOR DATA COLLECTION

Quantitative methods
Focus on numbers and frequencies “Numbers.”
- circulation, web usage analytics, survey data (not free text), gate counts, number of classes taught

Qualitative methods
Capture descriptive data and focus on experience and meaning. “Words.”
- Usability testing, focus groups, user interviews, ethnographic studies, observational studies
Existing data or new data?
CONSIDERATIONS FOR DATA COLLECTION

Before you begin: data requirements

➢ Know what questions the data needs to be able to answer
➢ Data structure requirements
➢ Data extraction capabilities
Effectively measuring change
operational data
“digital exhaust data”
transaction data
administrative data
Common evaluation methods

- Usability testing
- Web usage data
- A / B testing
- Surveys
- Focus groups
- Pre / post testing
- Cost/benefit analysis
Common evaluation methods

- Usability testing
- Web usage data
- A / B testing
- Surveys
- Focus groups
- Pre / post testing
- Cost/benefit analysis
A/B testing

- Involves an online performance comparison between a webpage control group and a single variable test.
What?

- Compare two potential workflows
- Research study to analyze differences in use rates for digital images that have received manual metadata enhancements versus images that have only minimal, collection-level metadata automatically extracted from the finding aid
How?

- One digital image collection
- A/B testing: half of the collection receives metadata enhancements by staff, the other half have only collection-level metadata
- Put online in the same interface, wait 6 months
- Google Analytics provides data to compare performance of our two test groups
Findings

- Images with manual metadata enhancements were used four times as frequently
- 92% of unenhanced images had still not been viewed even once after 6 months
- Enhanced images had been viewed at least once at a rate three times higher
- Person names were included in 28% of search strings that led to page views (person names were only available in enhanced metadata)
Google Analytics offers free tools for A/B testing
**Cost/benefit analysis**

- While we assume there to be inherent value in the work we do, libraries are almost completely lacking in metrics for measuring cost and value.

- Unlike for-profits, we cannot measure “cost” against “sales” – the traditional measure of value.
We must create our own operational definitions of value:

- Discovery success, use, display understanding, data’s ability to operate on the open web, throughput/timeliness, etc.
What?

- Cost/benefit analysis of quality control visual checks for large-scale digitization

- Cost =
  - Staff time to conduct visual checks
  - Opportunity cost (lost time towards production)

- Value =
  - The quantity, severity, and type of errors uncovered and corrected during visual checks
How?

- Collected time data for scanning and quality control over a 3-month period
- Tracked folder IDs for each QC batch, IDs linked to filesystem data about how many scans were in a folder
- Tracked error types in 6 categories, each tagged as “critical” or “non-critical” (depending on whether the error caused the user to be unable to read/use the item, or only caused inconvenience).
Findings

- 85% of time was spent scanning; 15% on quality control
- One error was discovered for every 223 scans (0.4%)
- Only 32% of all errors were “critical”
- There was one critical error for every 700 scans (0.1%)
Secondary findings: large folders

- Folders with 100+ scans = 11.5% of all folders
- 37% of folders in this group contained errors
- 30% of all errors occurred in this 11.5% of folders, and 52% of all critical errors occurred in these folders
- Performing visual checks on the large folders required 32% of all visual check time
Conclusions

- If all the time spent performing visual checks were instead spent on scanning, production would have increased by 18%
- Reviewing larger folders more frequently than small folders would increase “bang for the buck” in QC
  - It would also provide a higher rate of detection for critical errors than a simple percentage-based sampling of all folders
- If no QC was performed at all, there would only be a critical error in 0.1% of scanned material (1 per 700 scans)
Thank you!

Joyce Chapman
▶ Consultant for Communications & Data Analysis
▶ State Library of North Carolina
▶ 919.807.7421
▶ joyce.chapman@ncdcr.gov