JUS
JOURNAL OF USABILITY STUDIES

Issue 1, Vol. 1, November 2005, pp. 27-45

Towards the Design of Effective Formative Test Reports

Authors:

Mary Theofanos
National Institute of Standards and Technology
100 Bureau Drive
Gaithersburg, MD 20899 USA
mary.theofanos@nist.gov

Whitney Quesenbery
Whitney Interactive Design
78 Washington Avenue
High Bridge, NJ 08829 USA
whitneyq@wqusability.com

Abstract

Many usability practitioners conduct most of their usability evaluations to improve a product during its design and development. We call these "formative" evaluations to distinguish them from "summative" (validation) usability tests at the end of development.

A standard for reporting summative usability test results has been adopted by international standards organizations. But that standard is not intended for the broader range of techniques and business contexts in formative work. This paper reports on a new industry project to identify best practices in reports of formative usability evaluations.

The initial work focused on gathering examples of reports used in a variety of business contexts. We define elements in these reports and present some early guidelines on making design decisions for a formative report. These guidelines are based on considerations of the business context, the relationship between author and audience, the questions that the evaluation is trying to answer, and the techniques used in the evaluation.

Future work will continue to investigate industry practice and conduct evaluations of proposed guidelines or templates.

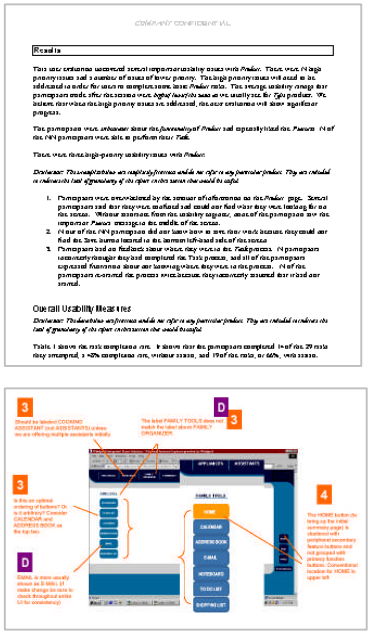


Figure 1. Sample reports included both formal reports and more graphical presentations. Even without reading the words, the difference in these two reports is visually striking. (Koyani and Hodgson)

Keywords
Information interfaces and presentation, usability testing, usability reporting, CIF, usability standards, documentation

Background

In 1998, the US National Institute of Standards and Technology (NIST) began the Industry Usability Reporting (IUSR) project to improve the usability of products by increasing the visibility of software usability. The more specific focus is to provide tools to clearly and effectively communicate the usability of a product to those considering the purchase of a product, and to provide managers and development teams with a consistent report format for comparing the results of usability tests.

Following the successful creation by IUSR of an international standard, the Common Industry Format (CIF) for summative usability test reports (ANSI-INCITS 354:2001 and ISO/IEC 25062: 2005) the IUSR project began a new effort to create guidelines for reporting formative usability test results.

The primary goal of this project is to help usability practitioners write reports of formative evaluations that communicate effectively and guide the improvement of products. The results of the project may include a library of sample reports, the creation of guidelines for report writing, templates that practitioners can use in creating their own reports, or even formal standards describing the process and content of formative usability test reports.

Investigative work in this project followed a consensus process similar to that of creating an industry standard, It focused on collecting industry report samples, comments and best practice guidelines from practitioners.

Method and Process

The formative reporting project has held two workshops to collect input from usability professionals from across industry, government and academia. The first was in Boston at Fidelity Investments in October 2004, with twenty-nine participants. The second was in Montreal at the UPA 2005 conference in June 2005, with nineteen participants.

For these workshops, participants provided sample formative reports or templates that were analyzed with respect to their form and content. In addition, the first workshop participants also completed a short questionnaire on the business context in which the reports were created. A detailed analysis of the reports and the questionnaires submitted for the first workshop was used as the basis for later work.

Four parallel efforts addressed the following questions, each discussed in this paper:

1. What are the possible elements of a formative report, and what are best practice guidelines for their use?
2. What are some of the ways in which key elements, especially observations, metrics, and recommendations are presented?
3. What is the role of metrics in reporting of formative usability tests?
4. How does the business environment influence report content or style?

Figure 2. The rules were gathered through an affinity technique. Participants first created cards with "if-then" conditions, then grouped them with other, similar rules. The initial set of 170 rules were reviewed, consolidated and re-categorized at the second workshop.

Separate breakout groups at the workshops focused on each of these areas. At the October workshop, one group created the super-set list of report elements, a second began collecting a set of "rules" (really, best practice guidelines, based on their practical professional experience) for making decisions about when to include an element in a report, a third examined the report context including author's context, and a fourth examined the use of metrics in formative reporting. At the UPA 2005 workshop, one group continued the work on metrics while the second focused on matching the proposed guidelines to the list of report elements and refining them.

This paper describes the current status of work in identifying report elements and the range of how these elements are presented in the sample reports. Some of the "rules," though still in progress, are included, identified with a check-mark style bullet.

Defining Formative Usability Testing

Our first task was to define formative testing and the scope of this project. The following definition was created at the first workshop and re-affirmed at the second.

> Formative testing: testing with representative users and representative tasks on a representative product where the testing is designed to guide the improvement of future iterations.

This definition includes any technique for working with users on any sort of prototype or product. These evaluations may be conducted in a lab, in an informal setting, or using remote access technology. It excludes heuristic reviews, team walk-throughs, or other usability evaluation techniques that do not include representative users. It also excludes surveys, focus groups, or other techniques in which the participants do not work with a prototype or product. It is possible that many of the best practices for formative usability reporting will apply to these other techniques, but that is not the focus of this project.

What are the elements of a formative report?

The workshop participants created a superset of elements that could be considered for a formative report would be a valuable starting point. After some light editing, this list includes some 88 common elements, grouped into 15 broad categories:

1. Title Page
2. Executive Summary
3. Teaching Usability
4. Business and Test Goals
5. Method and Methodology
6. Overall Test Environment
7. Participants
8. Tasks and Scenarios
9. Results and Recommendations
10. Detail of Recommendations
11. Metrics
12. Quotes, Screenshots and Video
13. Conclusions
14. Next Steps
15. Appendices

Perhaps it is predictable, since the motto of the usability professional is "it depends," but there was not one element that appeared in all of the sample reports. One proposed rule even pointed out that a report might not always be necessary.

- ✔ If no one will read it, don't write a report.

With that said, however, the most basic elements - some information about the participants, a description of the tasks or activities, and some form of results and recommendations - were the core of all of the reports.

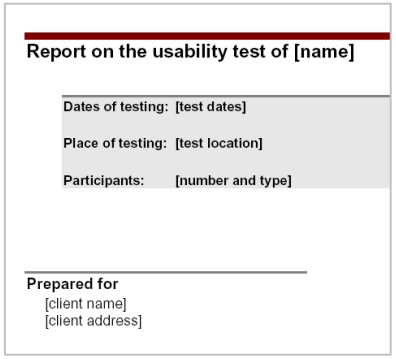


Figure 3. This report cover provides details about the test event – dates, location and type of participants - in a succinct format.. This sample was one of the clearest statements of these details, making them part of the cover page in a standard template. (Redish)

Some of the elements that appeared in few of the reports seemed a matter of the "professional style" of the author—or perhaps in response to the reporting context. For example, the use of a table of contents, the inclusion of detailed test materials in the report, or acknowledgements may all be a function of the difference between using a presentation or document format, rather than a disagreement about the value of those elements in some cases.

We analyzed the sample reports, identifying which elements had been included in each of them. The authors all reviewed this analysis to ensure that their intent was accurately understood. This was especially important for report *templates* and reports with sections of content removed for confidentiality. During this analysis, we also resolved some ambiguities in the element names, and removed some duplications. A table at the end of this paper lists all of the elements as they were presented at the second workshop.

This work revealed that only 25 elements of the 88 total common elements appear in at least half of the reports. Expanding the criteria to include elements in just a third of the reports brings the total to 39. We were surprised at this finding, given the general consensus around the list of elements. Is there a gap between our *beliefs* and our *practice?* Or are there variations in reporting styles that affect the content of the reports.

Elements: Introduction and Background

During discussions about reporting elements, there were many comments about the importance of including a lot of detail about the purpose and context of the report. The most common elements in this section are:

- **Title page information -** Basic identification elements include the author and date of the report and the product being tested. Interestingly, the date of the test and name of the testers was noted in only a small number of the reports.
- **Executive summary** - About half of the reports included a summary (though it is important to note that some of the reports were intended, in whole, for executives).
- **Business goals** - Over three-quarters of the samples listed the business goals for the product.
- **Test goals** – Two-thirds of the reports also defined the goals of the test.

Other elements in this group are a statement of the scope of the report and more detailed descriptions of the product or artifact being tested. Background information about usability was included in only eight percent of the reports, but appeared to be a standard element for those authors.

Why do usability testing?

- To observe representative practitioners and researchers using your website as they do in real life
- To document successes and challenges as they navigate the site
- To ensure the site is the best it can be:
  - Efficient
  - Effective
  - Satisfying
- To ready the site for future enhancements and content

Figure 4. Reports intended for an executive audience were the most likely to include a description of usability testing. These descriptions were often general, like this example. (Schumacher)

Elements: Test Method and Methodology and Overall Test Environment

It was a little surprising that only two-thirds of the reports mentioned the test procedure or methods used. Less than half described the test environment, and even fewer described the software (OS or browser type) or computer set up (screen resolution).

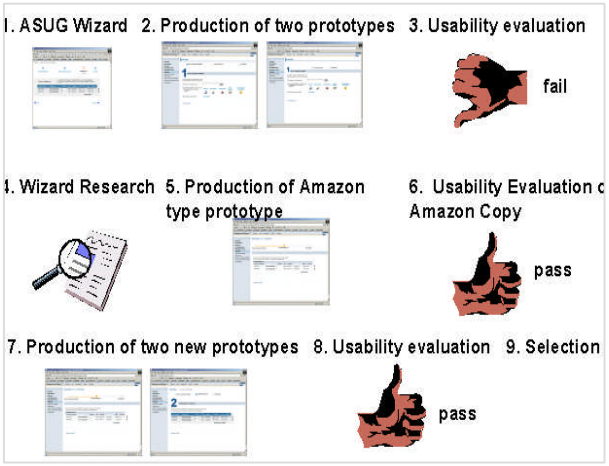


Figure 5. This report presented the history of an iterative prototype-and-test user-centered design process in a single slide. This approach sets a strong context for the rest of the report. (Corbett)

A few reports included information about usability or the user-centered process that the test was part of. The rules suggested for deciding whether to include this sort of information make it clear that this decision is entirely based on the audience and ensuring that they will be able to read the report effectively.

- If the audience is unfamiliar with usability testing, include more description of the method and what usability testing is.
- If the audience is unfamiliar with the method, include a brief overview.
- If the team observed the tests, don't include methodology.
- If this report is on one in a series of tests, be brief and don't repeat the entire methodology.

Elements: Participants

Almost every report included a description of the test participants. The level of detail ranged from a simple statement of the number and type of participants to more complete descriptions of the demographics.

The most common presentation was a simple chart listing the (presumably) most important characteristics for the test.

User	Gender	Age	Ethnicity	Income	Ed.	Computer	Home Type	Yard	Environ.
1	Female	55-64	Caucasian	40-60	BA	Basic	Townhouse	Yes	Medium
2	Female	25-34	Hispanic	<40	BA	Medium	Apartment	No	Low
3	Male	65+	Caucasian	Retired	MS	Basic	House	Yes	Medium
4	Male	35-44	Non-White	60+	BA	High	House	Yes	Low
5	Female	35-44	Caucasian	40-60	BA	Medium	Townhouse	Yes	Medium
6	Female	35-44	Russian	60+	MS	Medium	House	Yes	Low

Table 1: Participant Characteristics

Figure 6. This is a typical chart showing participant characteristics. Some information, such as the gender, age and ethnicity could be used for any report. Other columns, such as the income, computer use and type of home are related to the product being tested. (Salem)

One report compared the characteristics of the participants to a set of personas. Other reports used internal shorthand for groups of users but did not specifically relate the test participants to design personas.

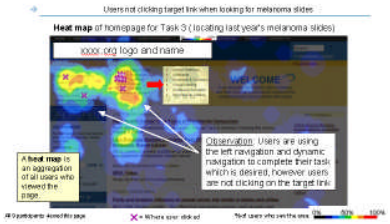


Figure 7. Some techniques, like eye-tracking, have strong visuals that can be used to organize findings. This report used callouts to identify patterns of behavior shown in the eye-tracking data and to relate it to usability problems. (Schumacher)

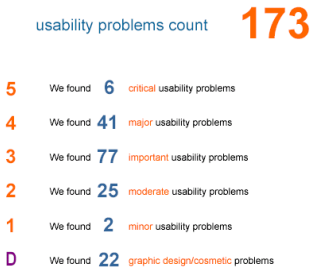


Figure 8. This report used the type and severity of problems to organize the information. It included this summary of problems found at each level along with an explanation of the levels. (Hodgson)

Elements: Tasks and Scenarios
Almost every report described the participant's tasks or scenarios, but this was another area where the level of detail and presentation style varied widely. At one extreme, tasks were defined only in a very general statement as part of the background of the report. At the other, the complete task description was given. In a few cases, the tasks were used to organize the entire report, with observations and recommendations presented along with each task.

Only a very few reports described success criteria, defined the level of difficulty of the task, or compared expected to actual paths through the interface.

Elements: Results and Recommendation, Detail of Recommendations and Quotes and Screenshots
All of the reports had some presentation of observations, problems found, and recommendations (though in varied combinations). Any consistency ends, however, at this statement. There were many effective formats within the sample reports, from presentations based almost entirely on screen shots and callouts to tables combining observations, definition of problems, recommendations for changes, and severity or priority ratings. The presentation of report findings is discussed in more detail in the next section.

Elements: Metrics
The most common metrics reported were the results of satisfaction questionnaires. In general, the use of metrics, from task success to time-on-task measures, seemed to be partly a question of professional opinion. Although there was general agreement on the value of reporting quantitative measures in many report contexts, some people seem to use them more consistently than others. Metrics are discussed in more detail later in this report.

Elements: Conclusions and Next Steps
The elements in this group are usually at a higher level than the more specific recommendations. They provide report authors with an opportunity for a more conceptual or business-oriented comments or to suggest next steps in the overall user-centered design process.

It was somewhat surprising how few (less than 20%) provided an explicit connection between the business or test goals and the results.

Although a third of the authors wrote a general discussion section, relatively few took advantage of this opportunity to recommend further work or to discuss the implications of the findings. It is not clear whether this is because those elements are seen as outside of the scope of a test report and are being done in another way. It may also depend on whether the report is considered a "strategic" document.

How are report findings presented?

As our definition of formative testing indicates, details of the test results—observations, descriptions of problems, and recommendations—are the heart of any usability report. They were also the elements of most interest to the project participants

We looked at both observations and recommendations, and we found a wide range of approaches to organizing and presenting them. In some cases, scenarios, observations, and recommendations are grouped

together; in others, they are presented as separate elements.

Are positive findings included in a report?

Only a third of reports included both positive findings and details of problems, but with varying degrees of focus. Positives were sometimes listed briefly near the beginning of the report; sometimes, they were mixed in with the list of issues or other findings.

	[[Feature Name]]
Good stuff	With minimal prompting, started using XXX to figure out red boxes and blue x's.
	Used the picture in XXX to find information
	Noticed the Status Bar when widening Notch filter
	Seemed to read XXX
	Bouncing around between XXX and Help viewer seemed transparent to her, "got it" that general stuff is in the Help browser
	Used WTC cursor with no problem
	Likes seeing status of various aspects (damping) in status bar

Figure 9. This report used a spreadsheet to summarize findings, including "good stuff," or positive findings. (Rettger)

Some reports had no mention of positives at all, but one proposed rule is that they should always be included. The most common reason cited was as a way to smooth any "ruffled feathers" and give encouragement. But they were also considered important as a way of documenting what did work, so it would not be changed.

- ✔ If persuasion is a reporting goal, focus even more than usual on positive findings.
- ✔ If the things you hoped would work actually worked, include positive findings.

How are findings organized and presented?

The findings were organized in many ways: by priority, topic, task/scenario, or screen/page. Professionals are clearly aware that the organization and presentation of the report, especially the core results, can have an impact on how well the report is received, and how persuasive it is.

- ✔ Issues of design impact are complex. Consider interactive, non-linear presentation of findings and recommendations.
- ✔ You were testing to answer specific questions, so report findings relevant to these questions.

Whether done in a document format, using word processing software, or in a presentation program, about half of the reports include screen shots or some other way of illustrating the results. Some were almost exclusively based on these screens.

- ✔ If the audience is upper management, use more visual elements (charts, screen shots) and fewer words.

There were strong sample reports that used each of these approaches. Many seemed to use a standard template, suggesting that the authors used the same approach in all of their reports. Others seemed more ad-hoc, possibly organized to match the usability testing technique used.

Some reports focused on presenting observations first. In one example, the report was organized by heuristics, with specific findings listed below, along with supporting quotes from users. Other reports omitted recommendations because the report is organized into several documents, each for a different group or part of the team process.

- ✔ If team responsibilities are well known, organize findings based on who is responsible for them.

Including more direct quotes or other material that brings the users' experience into the report was

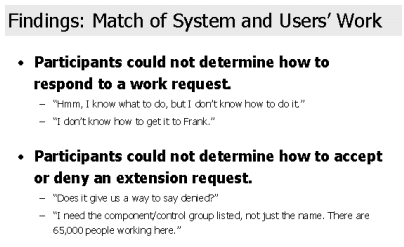


Figure 10. Direct quotes are an important feature of this report's observations, and used to support all findings. (Battle/Hanst)

considered a good way to help build awareness of user needs.

- If the audience is not "in touch" with users, include more subjective findings to build awareness (and empathy).
- Don't confuse participant comments with a "problem" that has behavioral consequences.

Reports which are organized by page or screen often used screen images, either with callouts or to illustrate the explanation of the observations.

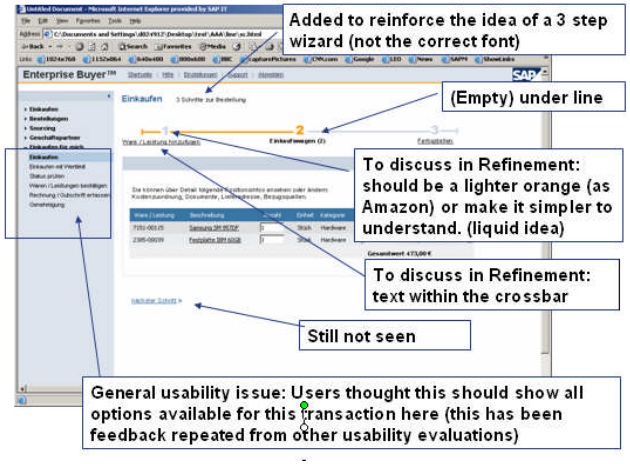
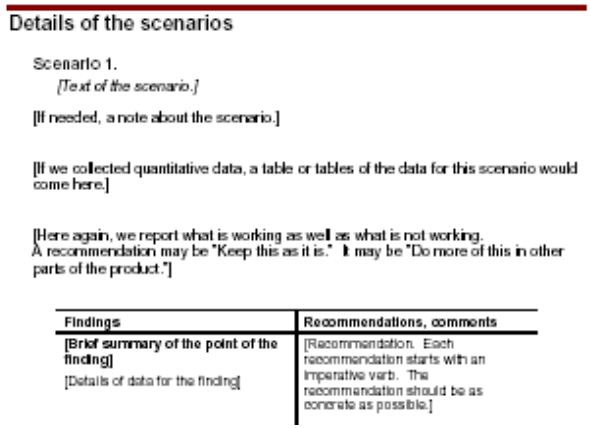


Figure 12. This is a typical example of a group of findings presented as callouts on an image of a page or screen. (Corbett)

This technique is also a good way to present minor issues, quick fixes, or other problems that were observed during the usability test.

- If you are reporting minor issues or quick fixes, organize them by page or screen.

Reports organized by task or scenario listed the scenario and sometimes the expected path along with observations. These reports often used a spreadsheet or table to create a visual organization of the material. but some simply listed the tasks and observations.



Details of the scenarios

Scenario 1.

[Text of the scenario.]

[If needed, a note about the scenario.]

[If we collected quantitative data, a table or tables of the data for this scenario would come here.]

[Here again, we report what is working as well as what is not working. A recommendation may be "Keep this as it is." It may be "Do more of this in other parts of the product."]

Findings	Recommendations, comments
[Brief summary of the point of the finding] [Details of data for the finding]	[Recommendation. Each recommendation starts with an imperative verb. The recommendation should be as concrete as possible.]

Figure 13. This template provides detail of findings by scenario. It describes the scenario, includes screen shots when helpful, and then presents the findings and recommendations, keeping all of this information together in one place. (Redish)

When the report will be used as the basis for discussion, there is an obvious value in keeping the scenario, observations, and problems together, as they are then presented in one place for easy reference.

- If you are using the results for discussion, include problem descriptions.

How are recommendations organized and presented?
The most important difference in how recommendations are presented is whether they are listed with each

finding or as a separate section in the report. One report did both, presenting recommendations as the conclusion to an observation, but also listing them all together at the end of the report.

Table 3: usability problems and recommendations

Usability problem	Recommendation	Severity
1 - Problem described in Full Detail	Recommended usability solution described in full detail here.	★★★★★
2 - Problem described in Full Detail	Recommended usability solution described in full detail here.	★★★★★
3 - Problem described in Full Detail	Recommended usability solution described in full detail here.	★★★☆☆

Figure 14. This table is organized by problem, presented in order of severity. Problems are described in a brief sentence, and each is accompanied by a recommendation. (Zelenchuk)

Another popular format grouped the statement of the observation, usability problem, and recommendation along with a screen shot illustrating the location of the problem.

A few reports did not include any recommendations, but in some of those cases, there were separate documents with either recommendations or detailed action lists. These were often reports created by a usability professional working within a team or with an ongoing relationship to the team. A few usability professionals used the initial report as the introduction to a work session in which the team itself would decide what actions to take. One rule suggested another reason to leave recommendations out:

- ✔ If there is a lot of distrust of the design/test team, emphasize findings and build consensus on them before making recommendations.

Severity Rating	How it is displayed	What it means
High	1 2 **3**	• Finding prevented participants from being able to successfully complete tasks. • Finding is a showstopper. • Findings rated as severe are the most critical issues and should be prioritized first.
Medium	1 **2** 3	• Participants were still able to accomplish tasks but had some difficulty in doing so. • Participants were frustrated or annoyed by the issue. • Participants see these issues as a nuisance.
Low	**1** 2 3	• Participants were able to use these features successfully with little to no difficulty.

Figure 15. This three-point scale is presented with an explanation for each rating. (Wolfson)

How strong are the recommendations?
The verbal style of recommendations varied from extremely deferential suggestions to strong, commanding sentences. This was partly based on how obvious a specific recommendation was. In some cases, there was a mix of styles within a single report. Authors used stronger language for more firm suggestions, and they used more qualified language when the direction of the solution was clear, but the specific design choices were not obvious.

The style of the recommendations was also influenced by a combination of writing style, the status of the usability professional, and business context.

- ✔ If you are a newer usability person, or new to the team, use less directive language and recommendations.

Are severity or priority levels included?
Severity or priority levels were another area that caused much debate. Some felt that they were critical to good methodology; others did not use formal severity ratings, but used the organization of the report to indicate which problems (or recommendations) were most important.

One of the rules suggests that priorities can be used to help align the usability test report with business goals:

- ✔ If executives are the primary audience, define severity, priority, and other codings in comparison to business goals or product/project goals.

When formal severity ratings were used, they were always explained within the document. Clearly, there is no industry consensus for rating problems. A workshop

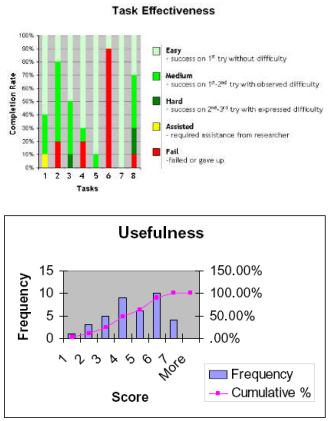


Figure 16. Quantitative metrics were often presented in visual chart formats, along with an explanation and supporting data. (Hodgson and Parush)

at UPA 2005 considered the problem of creating a standard severity scale.

How do metrics contribute to creating recommendations?

Reports tended to include–or leave out–whole groups of elements relating to metrics or detailed quantification of observations being reported. This was partly based, simply, on whether those measurements were taken during the test. Any disagreements about metrics were over whether they should be included, rather than over how to report them.

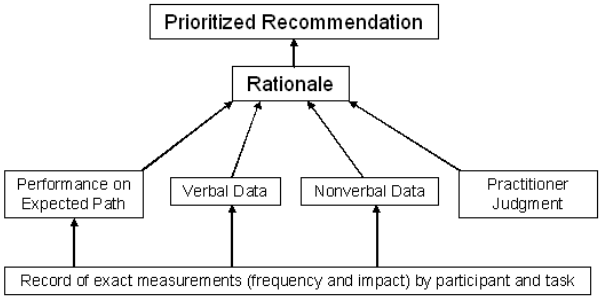


Figure 17. In a metric-centered view, the record of exact measurements feed analysis of performance, verbal data and non-verbal data. This is blended with the practitioner's judgment to form the rationale for making prioritized recommendations. (From UPA 2005 Workshop Report)

The group working on metrics took a very high-level view of the role of metrics. In their view, the measured details of the test session, blended with the practitioner's judgment, are the basis for making recommendations based on usability testing. There were others, however, who felt that the emphasis on "counting" was at odds with a more qualitative and empathetic approach.

What metrics were considered?

The following metrics were most often used in the sample reports or mentioned in the workshop discussions:

- Overall success—Is there a scoring or reporting of overall success of the user interface?
- Task completion—Does the report track successful task completion? Does it use a scale?
- Time metrics—Is there any timing information reported like time on task or overall time?
- Errors made—Does the report track errors and recovery from errors?
- Severity—Are problems reported assigned a severity?
- Satisfaction—Are user satisfaction statistics included and what survey was used?

By far the most common performance metric reported was task success or completion. This was generally measured for each task and was either binary, measuring whether the task was successfully completed or not, or measured as "levels" of completion. This data was most often presented as the percentage of users who successfully completed each task or the percentage of tasks each user successfully completed, or as averages across tasks or users.

Almost half the reports created metrics to quantify usability issues or problems with the system. These metrics were usually based on data analysis, not collected during each session. Examples included the number of issues, the number of participants

encountering each issue, and the severity of the issues. Some also included an overall rating summarizing the results of the test.

Time on task was less commonly reported. This was typically measured as elapsed time from start to finish of each task. It was also presented as the average time per task across users. Even fewer reports included performance metrics such as the number of errors and number of clicks, pages, or steps.

Understanding the reporting context

The need to understand the report context was a pervasive theme in almost all of the discussions and activities. It became clear that the decision to include an element and how to present the element was based on context. In summative reporting, the Common Industry Format (CIF) standard focused on a consistent presentation of a common set of information. In formative reporting, on the other hand, the author needs more leeway to tailor the report for both the specific audience and general context, including:

- **Who** is the author, and who is the audience for the report?
- **What** is the product being tested?
- **When** in the overall process of design and development is this report being created?
- **Where** does the report fit into the business and its approach to product design and development?
- **Why** was this usability test conducted, and what specific questions is it trying to answer?

This need to understand the context in which the report is created is critical to the development of a formative report.

Does business environment influence report content or style?

One of our early questions was whether the business environment had an influence on report formats. For the Boston workshop, participants answered a questionnaire that covered:

- the industry in which the report was used
- the size of the company
- the type of products evaluated
- the product development environment
- the audiences for the report
- the usability team and its relationship to the product team and the audience
- whether there was a formal usability process in place

Although these questions produced some interesting qualitative data, we could find no strong correlation between the content or formality of the report and the business setting. As importantly, different styles of reports or approaches to designing a report did not mean that substantially different content was included, though the level of detail varied widely.

We did however, find some patterns in the audiences for a report and the relationship between author and audience.

Relationships and goals are an important starting point
As the authors, we place ourselves in the center of the picture, and focus on our own goals or relationships as part of the overall goal in creating a usability report.

Five key relationships between the report author and readers were frequently mentioned, with different implications for the style of the report.

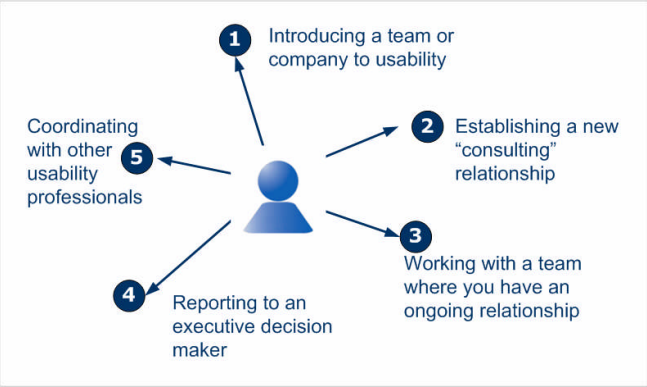


Figure 18. These relationships are seen from the perspective of the usability professional writing the report.

1. Introducing a team or company to usability
   Strong focus on teaching the process and explaining terminology.
2. Establishing a new "consulting" relationship
   Need to establish credibility or fit into an existing methodology.
3. Working within an ongoing relationship
   The report may be one of a series, with a style and format well known to the audience.
4. Reporting to an executive decision maker
   Need to draw broader implications for the test findings, as well as ensuring that the report is cast in business terms.
5. Coordinating with other usability professionals
   Include more details of methodology and relationship to other usability work.

There are many audiences for a report
There were four basic audiences consistently identified for the usability reports and a related focus for the report:

- **Executive Management**
  Focused on the business value of the work
- **Product Design/Development Team**
  Creating recommendations and actions
- **Usability or User-Centered Design Team**
  Communicating the results of the test
- **Historical Archives**
  Preserving the results for future reference

Several of the rules remind practitioners to consider the audience:

- ✔ Always identify the audiences who will read the report.
- ✔ If there are multiple audiences, segment the report to address each of them.
- ✔ If the organization's culture has certain expectations, identify them and try to meet them.
- ✔ If you are delivering to managers or developers, make it as short as possible.

Not all of the participants made a distinction between the product team and the usability team. This seemed to depend primarily on whether the report was written

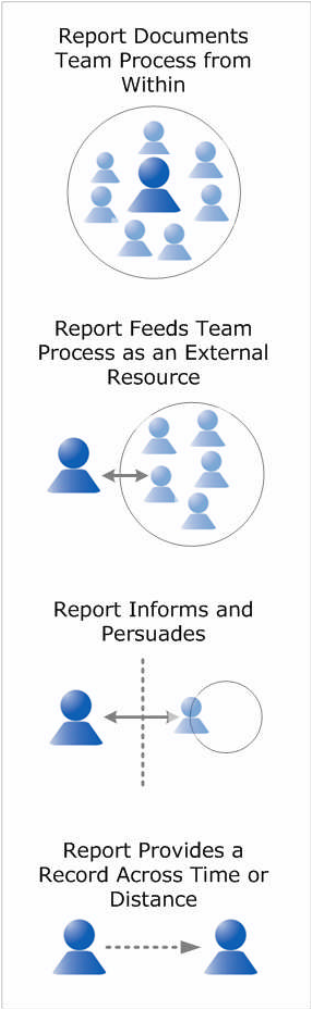


Figure 19. The relationship of the author to the team interacts with the role of the report.

by someone who was a permanent part of a project team, or by a "consultant," whether an internal resource (from a usability department within a company) or external consultant (someone engaged from outside of the company). Not surprisingly, in companies with several usability teams, report authors tended to be more aware of their usability colleagues as an audience.

Reports fill different roles in a team's process

One way to look at the relationship between the author and readers of the report is to consider the role that the report must play and the possible business relationships the author may have with the team and product management.

- When the report documents a team process, the usability report author is often part of the team and is working within an established methodology.
- If the author is related to, but not part of, the team, the report can feed the team's process, while providing an external perspective.
- A report must often inform and persuade both those within a team and executive management reviewing the work.
- When usability professionals must coordinate, whether over time (for example, building on previous work), or between projects. the report may have to serve an archival purpose, preserving more detail of the methodology and other details than in other contexts.

A report may be delivered in several stages or formats over time.

Several of the participants supplied more than one format for reporting on a single usability test. In some cases, these were done at different times following the usability test. For example, one participant said that she delivered an "instant report" by email, within a day of the test. She would bring back recommendations for work in the form of checklists, but then followed up with a more complete report within two weeks.

What are the implications of this work so far?

There was one other goal for reporting that we heard often: that reports should be short. Despite universal nods of agreement, the samples ranged from a 5-page template to a 55-page full report, with average length of 19 pages. Clearly there is a conflict between our wish to be concise and a need to explain the results of a test clearly and completely.

The wish for "short" reports may really be a wish for reports that "read quickly." It suggests a need to increase our information design skills, so we can present information in a format that can be scanned quickly. Our goal in this early work on designing effective formative test reports was to look for shared practices in the professional community, identifying the underlying logic of similarity and differences.

Throughout this project, so far, we have found both many similarities and differences. Despite debates over some aspects of methodology, there was a general consensus on the need for reports to address their specific audience and to make a persuasive case for the results of the test. Within this common goal, however, there were many differences in the details of what was reported and how it was presented. There are two main

sources of these differences, both rooted in professional experience and identity.

The details of the professional and business context of each report were very important to the workshop participants. Decisions on reporting format or content were usually defended with a story or explanation rather than with methodology. The design of a report is not simply a reflection of the test method, or even simple business needs. Even the "If…then…" format of the rules generated at these workshops reflects the contextual nature of these decisions. They also reflect the desire to encapsulate and communicate complex professional experience in a usable format.

The second source of variation in report formats is more personal. As each report author struggles to find an ideal way to communicate, their own taste and skill in information design is an important force. They may be trying to establish a unique authorial voice, or to fit seamlessly into a corporate culture – whether they are part of that company or an external consultant.

Both of these forces suggest that there cannot be a single "ideal" template for formative usability test reporting. Instead, we need best practice guidelines that address core concepts of user-centered communication: understanding the audience and presenting information in a clear, readable manner. This will require encapsulating good professional judgment in useful "rules" and accepting a wide degree of variation in presentation styles.

This approach is harder than a single, prescriptive standard template, but may be more beneficial to the development of the profession. It will encourage practitioners to consider how, when, and in what format they report on formative usability testing to be as important as the tests themselves. In the end, the goal of formative usability testing is to "guide the improvement of future iterations (of a product)." We cannot meet this goal unless we can communicate our findings and recommendations deftly – with art and skill.

Next Steps

Formative evaluations are a large component of usability practice yet there is no consensus on guidelines for reporting formative results. In fact, there is little literature on how to *report on* usability evaluations, in contrast to the many articles on methods for *conducting* evaluations.

One problem is that reports are often considered proprietary and rarely shared among usability professionals. Through these workshops, we were able to collect and analyze formative reporting practices. In fact participants agreed that one of the most valuable aspects of the effort was the ability to see what, how, and why others reported.

We found that there is more variation based on context in formative reports than summative reports, and thus it may not be practical to develop a standard format as a companion to the CIF. However, the high level of commonalities identified suggests that the development of guidelines and best practices based on business context would be valuable for usability practitioners.

As the project continues, we plan to continue gathering best practices from practitioners as well as conducing research with audiences who read usability reports.

Short term, we hope to publish a collection of sample reports, and create templates that organize the most often-used report elements into a simple format useful for new practitioners. Longer term, we will continue our work to generate guidelines for creating reports, and using metrics in reporting formative usability results.

Ultimately, we hope to provide practitioners with a better understanding of how to create reports that communicate effectively and thus exert a positive influence on the usability of many kinds of products.

Practitioners' Takeaways

- There is little guidance in the literature for the good design of a report on a formative evaluation. This is in contrast to summative evaluation reports, for which there is an international standard.
- There is wide variation in reporting on formative usability evaluations.
- The audience for a formative usability report should be carefully considered in designing the report format. The content, presentation or writing style, and level of detail can all be affected by differences in business context, evaluation method, and the relationship of the author to the audience.
- The IUSR project seeks to provide tools such as best practice guidelines and sample templates to help practitioners communicate formative evaluation results more effectively. To join the IUSR community, visit http://www.nist.gov/iusr/

Appendix: Common Elements in Formative Usability Reports

This list of elements was first created at the workshop in Boston. It was consolidated and categorized, based on David Dayton's analysis, and presented at the UPA 2005 workshop in June 2005.

Twenty-four reports were examined in detail. The numbers in the last column show percentage of these reports that included each element.

Title page and global report elements			
E1	Title page or area	71	%
E2	Report author	71	%
E3	Testers names	4	%
E4	Date of test	21	%
E5	Date of report	71	%
E6	Artifact or product name (version ID)	88	%
E7	Table of contents	58	%
E8	Disclaimer	0	
E9	Copyright/confidentiality	13	%
E10	Acknowledgements	0	
E11	Global page header or footer	67	%
Executive Summary			
E12	Executive Summary	50	%
Teaching Usability			
E13	How to read the report	8	%
E14	Teaching of usability	8	%
Introduction			
E15	Scope of Report	38	%
E16	Description of artifact or products	29	%
E17	Business goals and project	71	%
E18	Test goals	67	%
E19	Assumptions	0	
E20	Prior test report summary	4	%
E21	Citations of prior/market research	8	%
Method and Methodology			
E22	Test procedure or methods used	67	%
E23	Test protocol (including prompts)	17	%
E24	Scripts	0	
E25	Limitations of study or exclusions	17	%
E26	Data analysis description	13	%
Test Environment			
E27	Overall Test Environment	42	%
E28	Software test environment	21	%
E29	Screen resolution or other details	21	%
E30	Data collection and reliability controls	4	%
E31	Testers and roles	8	%
Participants			
E32	List or summary of participants	92	%
E33	Number of participants	83	%
E34	Demographics or specific background	46	%
E35	Relevance to scenarios	38	%
E36	Experience with product	29	%
E37	Company, employee data	0	
E38	Educational background	21	%
E39	Description of work	21	%
E40	Screener	17	%
Tasks and Scenarios			
E41	Tasks	67	%
E42	User-articulated tasks	4	%
E43	Scenarios	46	%
E44	Success criteria	13	%
E45	Difficulty	0	
E46	Anticipated paths	8	%
E47	Persons on the task	0	

Results and Recommendations			
E48	Summary	63	%
E49	Positive findings	67	%
E50	Table of observations or findings	58	%
E51	Problems / Findings	88	%
E52	Recommendations	83	%
E53	Definitions of coding schemes	17	%
Detail of Recommendations			
E54	Severity of errors	25	%
E55	Priority	13	%
E56	Level of confidence	4	%
E57	Global vs specific	13	%
E58	Classification as objective and subjective	29	%
E59	Reference to previous tests	4	%
E60	Bug or reference number	4	%
E61	Fix - description and effort	38	%
Metrics			
E62	Performance data / summary	38	%
E63	Satisfaction questionnaire results	50	%
Quotes, Screenshots and Video			
E64	Quotes	42	%
E65	Screenshots with callouts	42	%
E66	Video clips	4	%
E67	Voice-over	0	
Conclusions			
E68	Discussion	33	%
E69	Interpretations or implications of findings	17	%
E70	Tie-back to test or business goals	13	%
E71	Lessons learned (test process, product)	8	%
Next Steps			
E72	Further studies recommended	21	%
E73	List of what needs to be done	8	%
E74	Ownership of issues	0	
E75	New requirements and enhancements	4	%
E76	Style guide updates	0	
Appendices			
E77	Test materials	42	%
E78	Detailed test protocol or scripts	0	
E79	Tasks and/or Scenarios	13	%
E80	Consent Forms	0	
E81	NDA	0	
E82	Preliminary Report	0	
E83	Data analysis	13	%
E84	Raw data	4	%
E85	Data repository	0	
E86	Data retention policy	0	
E87	Statistics	13	%
E88	Style guide (for product)	0	

Acknowledgements

Our special thanks to workshop session leaders. This paper draws from material presented or created at the two workshops:

- Ginny Redish presented the first analysis of results and recommendations.
- Tom Tullis led the first workshop session on metrics; Jim Lewis, the second.
- Anna Wichansky led the session that created the initial list of elements. It was edited and organized by Elizabeth Hanst and David Dayton.
- Carolyn Snyder came up with the original idea for the rules and led the sessions on generating rules in both workshops.
- David Dayton did analyzed the sample reports, mapping how the elements are used in them.

The participants in the workshops were:
Bill Albert, Bob Bailey, Dean Barker, Carol Barnum, Lisa Battle, Nigel Bevan, Anne Binhack, Joelle Carignan, Stephen Corbett, David Dayton, Duane Degler, Cindy Fournier, Catherine Gaddy, Elizabeth Hanst, Haim Hirsch, Philip Hodgson, Ekaterina Ivkova, Caroline Jarrett, John Karat, Clare Kibler, Sanjay Koyani, Sharon Laskowski, James R. Lewis, Sara Mastro, Naouel Moha, Emile Morse, Janice Nall, Amanda Nance, Katie Parmeson Avi Parush, Whitney Quesenbery, Janice (Ginny) Redish, Mary Beth Rettger, Michael Robbins, Anita Salem, Jeff Sauro, Bob Schumacher, Carolyn Snyder, Mary Theofanos, Anna Wichansky, Janet Wise, Cari Wolfson, Alex Zaldivar, and Todd Zazelenchuk.

References

ANSI-NCITS 354-2001.(2001) Common Industry Format for Summative Usability Test Reports

IUSR Community Web site: www.nist.gov/iusr

Theofanos, M., Quesenbery, W., Snyder, C., Dayton, D. and Lewis, J. (2005). Reporting on Formative Testing: A UPA 2005 Workshop Report. Usability Professionals Association. http://www.usabilityprofessionals.org/usability_resources/conference/2005/formative%20reporting-upa2005.pdf

Further Reading on Usability Reporting

Butler, K. ; Wishansky, A.; Laskowski, S. J.; Morse, E. L.; Scholtz, J. C (2002) Quantifying Usability: The Industry Usability Reporting Project Human Factors and Ergonomics Society Conference July 01, 2002

Jarrett, C. (2004) Better Reports: How To Communicate The Results Of Usability Testing Proceedings of STC 51st Annual Conference, Society for Technical Communication, Baltimore, MD., May 9-12, 2004

Scholtz, J. C. ; Wichansky, A.; Butler, K. ; Morse, E. L.; Laskowski, S. J. (2002) Quantifying Usability: The Industry Usability Reporting Project. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Baltimore, Sept 30-Oct 4, 2002

Yeats, D., Carter, L. (2005) The Role of the Highlights Video in Usability Testing: Rhetorical and Generic Expectations. Technical Communication, Volume 52, Number 2, May 2005, pp. 156-162(7)

Authors

Mary Frances Theofanos is a Computer Scientist in the Visualization and Usability Group in the Information Access Division of the National Institute of Standards and Technology. She leads the Industry Usability Reporting (IUSR) Project, developing standards for usability such as the Common Industry Format for Usability Test Reports. Previously, she was the Manager of the National Cancer Institute's (NCI) Communication Technologies Research Center (CTRC) a state-of-the-art usability testing facility. Before joining NCI, Mary spent 15 years as a program manager for software technology at the Oak Ridge National Laboratory complex of the U.S. Department of Energy.

Whitney Quesenbery is a user interface designer, design process consultant, and usability specialist with a passion for clear communication. As the principal consultant for Whitney Interactive Design, she works on projects for large and small companies around the world. Previously, she was a principal at Cognetics Corporation. Whitney co-chaired the 6th NIST IUSR Formative Testing Workshop and is chairperson of the IUSR working group on formative testing. She is President of the Usability Professionals' Association (UPA), a past-manager of the STC Usability SIG, and Director of Usability and Accessibility for Design for Democracy (a project of UPA and AIGA).