

Reporting on Formative Testing

A UPA 2005 Workshop Report

*Mary Theofanos
Whitney Quesenbery
Carolyn Snyder
David Dayton
Jim Lewis*

August 8, 2005



1	INTRODUCTION	3
	Background to the workshop	3
	Aims of the workshop	3
	People involved	4
2	PREPARATION AND INTRODUCTIONS	5
	Submission process	5
	Introduction to the workshop	5
	Context of use for reports	5
	Common elements in formative reports	8
	Rules for determining the elements of a report	9
	Key types of metrics	9
3	BREAKOUT SESSIONS	10
	Metrics Breakout Group	10
	Task-Level Measurements.	10
	Event-Level Measurements.	11
	Back to Basics	11
	Rules Breakout Group	13
	How we worked on the rules	13
	Rules for general characteristics of the formative report	13
	Rules for positive findings	14
	Rules for detailed findings and problems	14
4	NEXT STEPS	17
	Common Elements and Rules	17
	Reporting Context	17
	Metrics Group	17
5	APPENDIX: RELATED MATERIALS	18
	Questionnaire on Context and Details of the Reports	18
	Common Elements in Formative Reports	19
	Proposed Rules for Reporting Formative Usability Results	22
	Formative Reporting Samples	32

1 Introduction

Background to the workshop

This workshop was a follow-up to the October 18-19, 2004 workshop on reporting formative test results held in Boston and led by the National Institute of Standards and Technology (NIST) Industry Usability Reporting (IUSR) project. At that workshop participants agreed on a working definition for formative testing: “Formative testing is a usability evaluation with representative users, using realistic tasks, on a representation of the “thing” being evaluated (e.g. a prototype or semi-working model) where the primary purpose of the evaluation is to guide the improvement in design of future iterations.” Participants focused on identifying:

- a superset of elements that may be included in formative reports,
- typical contexts of use for reports,
- different personas for developing reports,
- different metrics appropriate for reports,
- rules to help determine which elements to include in reports.

The goal of this second workshop held at UPA 2005 in Montreal, Quebec, was to review the work in progress and take the work to the next step by mapping the personas, context, metrics and rules to the superset of elements in order to develop a template, guidelines and best practices for reporting formative test results.

Aims of the workshop

The Common Industry Format (CIF) for usability test reporting of summative testing has been an ANSI/INCITS Standard since late 2001. However, organizations and usability practitioners are modifying the CIF to use it for reporting formative testing. This suggests an “unmet need” that the NIST Industry Usability Reporting (IUSR) project participants have identified as an area to investigate.

For many practitioners, most usability evaluation is formative, aimed at improving a product during its design and development. With many more techniques available for this type of work, practitioners need clear definitions goals and guidelines to use in planning and conducting formative evaluations, especially in communicating/reporting the work (and its value) to colleagues and customers.

The goal of this workshop was to review the work begun in Boston in October 2004 and to take it to the next level by mapping the personas, context, metrics and rules to the superset of elements in order to develop a template, guidelines and best practices for reporting formative test results.

People involved

The organizers of this workshop are the NIST project leaders for the IUSR (Industry Usability Reporting) community. Mary Theofanos is also managing the fast-track of the Common Industry Format (CIF) for Usability Test Reports as an ISO standard.

The following participants attended the workshop.

Organizers and Facilitators:

Name	Company
Whitney Quesenberg	Whitney Interactive Design, LLC
Mary Theofanos	NIST
James R. Lewis	IBM
Carolyn Snyder	Snyder Consulting
David Dayton	Towson University

Participants:

Name	Company
Bill Albert	Fidelity Investments
Dean Barker	Human Factors International
Anne Binhack	Yahoo
Joelle Carignan	SAP Labs
Cindy Fournier	Merge Interactive
Haim Hirsch	Kaiser Permanente
Sharon Laskowski	NIST
Sara Mastro	Weatherbug
Amanda Nance	Georgia Tech
Katie Parmeson	Landmark Graphics
Avi Parush	Carleton University
Ginny Redish	Redish & Associates, Inc.
Jeff Sauro	Oracle
Bob Schumacher	User Centric, Inc.

2 Preparation and Introductions

Submission process

Each participant submitted a sample formative report or a formative report template and completed a questionnaire describing the context and details of the sample report.

The questionnaire is included in the appendix.

Introduction to the workshop

At the NIST workshop in October 2004, the group agreed on a definition of formative testing to be used for this project. This definition was provided and accepted by the UPA 2005 workshop participants.

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.

The workshop started with a brief background and history of the IUSR project presented by Mary Theofanos and an overview of the submitted reports from the first workshop.

We then reviewed the 5 products developed at the NIST October workshop and work done to develop them since that workshop

1. the context of use for reports, and “personas” for report authors,
2. a superset of elements that may be included in formative reports,
3. a set of rules to determine the elements of reports, and
4. the different metrics appropriate for reports.

Context of use for reports

Whitney Quesenbery presented three different views of the context for formative reporting. In the first workshop, we found that each of these views influences the reporting format.

- Five critical business context in which the report can be delivered
- Four relationships between author and readers
- A view of reporting over time

Contexts for formative usability reporting



Five critical contexts

1. Introducing a team to usability
2. Establishing a new “consulting” relationship
3. Working with a team where you have an ongoing relationship
4. Reporting to an executive decision maker
5. Coordinating with other usability professionals

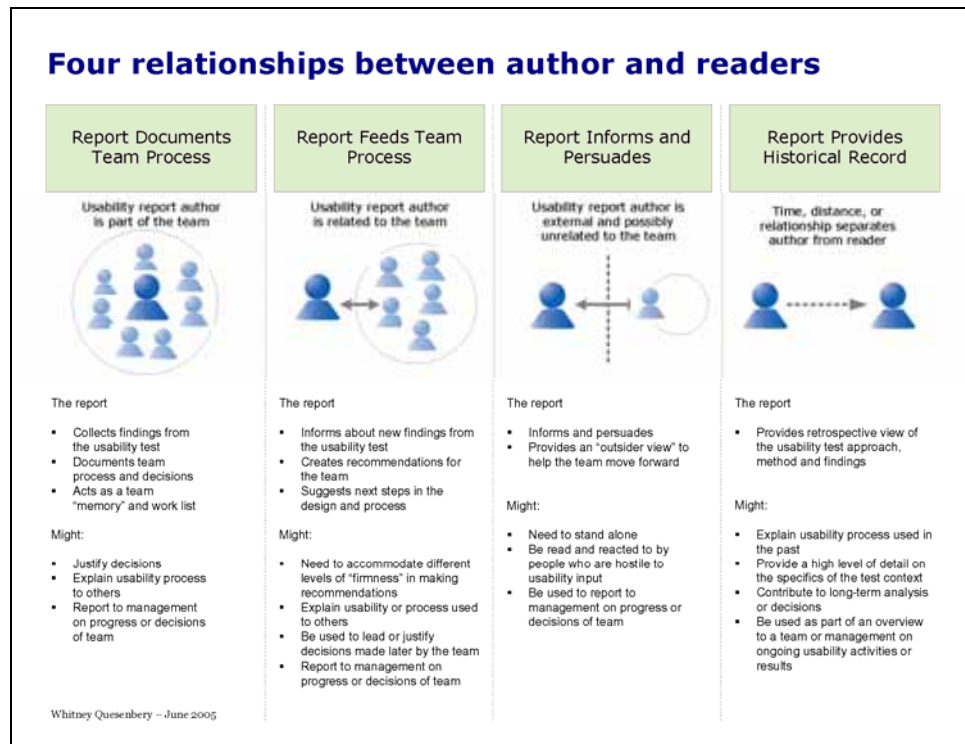
The five critical contexts look at the business or team relationship between the usability professionals and the people the report is delivered to. They are not (for the most part) mutually exclusive: several contexts may co-exist for any report.

- Each of the contexts creates concerns that the author must take into consideration. For example:
- Introducing a team to usability creates a need to explain the approach, metrics and results carefully, as well as being careful about terminology
- Establishing a new relationship suggests a need to establish credibility as you meet a team for the first time, as well as the possibility of needing to fit into an existing methodology or approach.
- Working within an ongoing relationship implies that you are part of the team either on a full-time or consulting basis, and that you work on a series of usability tests or related work.
- Reporting to an executive decision maker often means that the report gets special attention in some way. It also might mean reporting to someone who is not used to reading usability reports.
- Coordinating with usability professionals is an emerging need, as there are more opportunities to share results across projects or over time; distributed teams have special coordination needs. It might also mean anticipating the needs of someone coming into the middle of a project and reading reports from previous usability work

Relationship between author and readers

This view overlaps with the five contexts, but looks at the relationship between the report authors and readers. It also looks at some of the roles the report might fill, and the implications for what the report might need to include.

- **Report documents a team’s process.** The author is part of the team.
- **Report feeds team process.** The author is related to the team, but not an ongoing team member
- **Report informs and persuades.** The author is external to the team (or company), and might not even be related to the tema
- **Report provides an historical record.** A distance of time or relationship separates the author from the reader



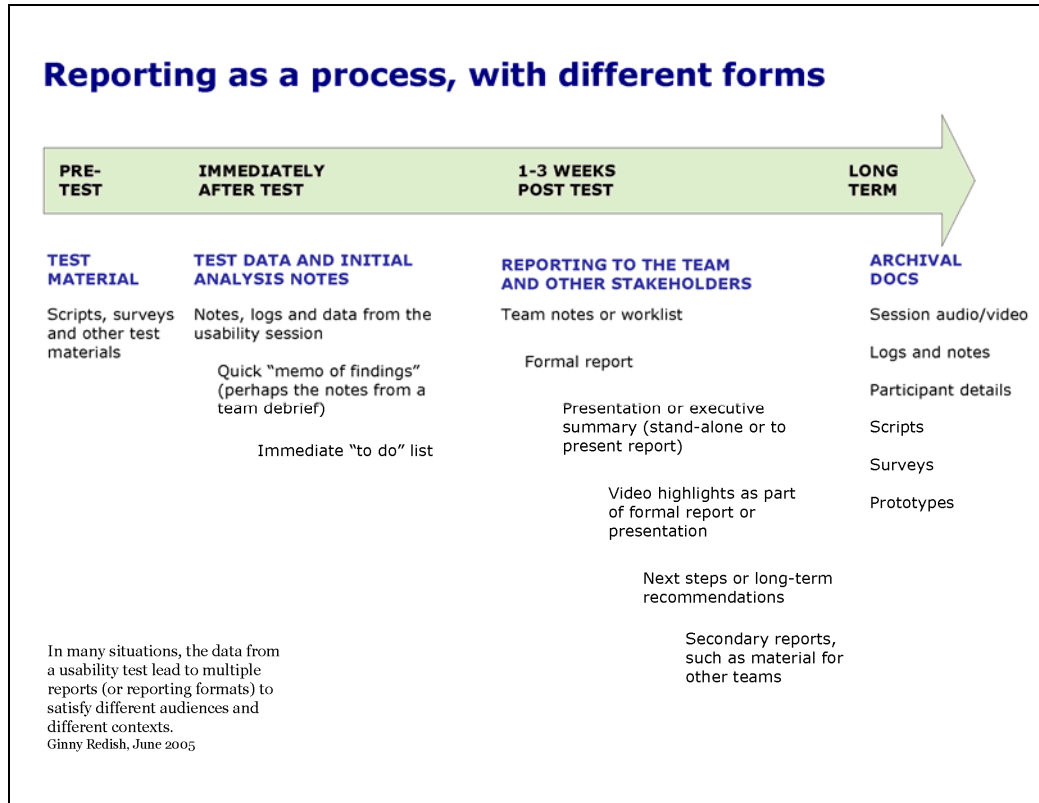
A view of reporting over time

One strongly stated view in the October workshop was that very few professionals delivered just a single report. Some provided both an “instant analysis” report, followed by a longer and more complete document. Others created different reports for different audiences. This led to a view of reporting as a series of events over time, suggested by Ginny Redish.

There are four time periods, each with distinct reporting needs:

REPORTING ON FORMATIVE USABILITY TESTING

- **Pre-test:** Test material, including scripts, surveys and other documents
- **Immediately after the test:** Test data and initial analysis notes, such as a quick “memo of findings” and an immediate to-do list
- **1 to 3 weeks post test:** Reporting to the team and other stakeholders. These reports include the “formal report,” but also video highlights, next steps or long-term recommendations, and any secondary reports (such as material for other teams)
- **Long term:** Archival documents including all session materials, audio or video tapes, scripts, prototypes, participant details and permission forms.



Common elements in formative reports

The workshop in October 2004 proposed a super-set list of 122 possible elements used in formative reports.

David Dayton reviewed the list of elements and the frequency with which they appeared in the submitted reports, mapping the content of the reports to the common elements. Based on this analysis, the original list was consolidated to 88 elements and organized into 15 groups. This consolidation eliminated duplicate entries, clarified ambiguous labels and eliminated some items that were explanatory rather than unique elements.

The list of elements is included in the appendix.

Rules for determining the elements of a report

At the October workshop, the participants developed a large collection of if/then rules, describing the conditions for including information in a report or deciding how to present it. An initial affinity analysis grouped similar rules together. Carolyn Snyder provided an introduction to the rules and how they relate (or not) to the common elements.

Before this workshop, the set of rules was re-analysed and group based on key decision points in preparing a report. This list of classified rules is included in the appendix.

Key types of metrics

Finally, Jim Lewis provided an overview of the key types of metrics including task-based scenario measurements including user-satisfaction scores, and problem/event based measurements.

This material was provided in a separate 28 page document, and is not included in this report.

3 Breakout Sessions

There were 2 breakout groups:

- the metrics group, led by Jim Lewis
- the rules associated with the common elements, led by Carolyn Snyder.

In addition during the course of the day Whitney interviewed seven (randomly selected) participants individually, gathering qualitative data on the contexts of the submitted reports and the roles of the authors. These interviews have not yet been transcribed or analysed.

Metrics Breakout Group

We started with a review of the work done at the workshop. At that workshop, the primary focus was on the use and reporting of task-level measurements, in particular, completion rates, completion times, and satisfaction. The point was raised that we should not preclude the reporting of other task-level metrics, such as number of assists, if the practitioner finds this data useful in developing recommendations. The point was also raised that not all formative tasks lend themselves to the collection of 'standard' task-level measurements and, in that case, such measurements should certainly not be reported.

The metrics group discussed task-level measurements and event-level measurements, and then switched gears to go back to basic questions.

Task-Level Measurements.

We focused on the "Big Three" (completion rates, completion times, satisfaction), but the point was also raised that practitioners should feel free to include any other task-level measurements (such as number of assists or number of clicks) that they would find useful (especially, useful as supporting data for problem identification or prioritization or as measurements to track over a series of formative studies to use to judge the incremental improvement in the product). We do not want to block the creativity of practitioners by mandating a strict set of task-level measurements.

Completion rates. We concluded that it is appropriate to report completion rates when the task lends itself to a pass/fail classification. Later discussion in the larger group brought up the point that even if the task lends itself to such classification, a practitioner might decide not to report it if there were reasons to believe that the recipient of the report would misuse the data (for example, would use 100% successful completion as a reason to avoid fixing problems that were serious, but not at the level of preventing task completion).

Completion times. There was some discussion of the usefulness of completion times in formative usability testing. On this point there was also considerable variation in practice. Some practitioners have no use for completion times that can be affected by methods such as think aloud, immediate interrogation of participants, and planned cutoff times for tasks. Other practitioners see value in collecting a

measure that is essentially free, as long as it has potential future use in later evaluations in a series of formative tests and has appropriate qualification in the report (for example, a statement that the time is unlikely to exactly predict performance with the final product).

Satisfaction. We did not discuss task-level satisfaction metrics to the same extent as completion rates and times. There did appear to be a consensus among practitioners that there is usually value in collecting such measurements. From the previous meeting, it appears that the basic technique is to have participants answer a small number of questions (such as those used in the ASQ). A potential enhancement is to also get similar data regarding the participant's expectations before completing the task. We did not discuss different means of analyzing and presenting this type of data, so this will have to wait for a future meeting or other form of discussion. We expect there will be considerable variation in practice.

Event-Level Measurements.

In formative testing, task-level measurements are of interest, but the guidance for future improvement of the product is more directly related to event-level measurements, where the events of interest are the things that happen (problems or positive events) as participants use the product under study. In practice, the focus is more on negative events (usability problems) than on positive events, but when there is expected value associated with the reporting of positive events (for example, easing the path to the discussion of problems, or ensuring that changes made to address problems do not result in the loss of positive events), then such events should be reported.

Event descriptions. A complete description of an event should include (1) a description of the event – what happened and, if possible, why it happened (it's possible that there might be a set of 'why' hypotheses – including hypotheses regarding whether the problem is specific or global), (2) the frequency of occurrence of the event, and (3) the impact on users.

Event analyses. There are several analytic processes applied against events and their associated metrics. One important one is to prioritize the events. We briefly discussed various published strategies for combining event-level metrics for the purpose of prioritization, but did not have time to discuss this in detail (note that there is detail about these schemes in the position paper distributed before the meeting). We do not appear to be in a position to strongly recommend the use of any one particular scheme. For formative reports, it seems reasonable that if a practitioner uses a systematic scheme for prioritizing events, then the report should include a description of that scheme.

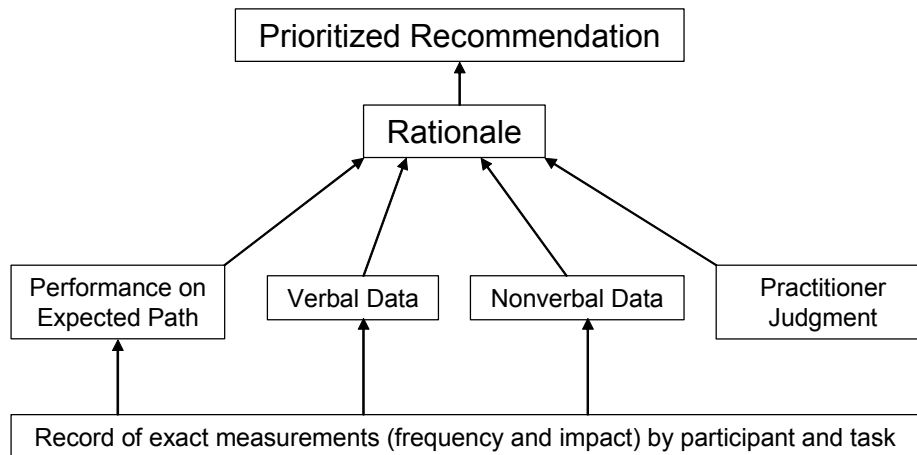
Back to Basics

At this point, we decided to change gears and return to a discussion of the basic purpose for conducting formative usability tests – which is to gather information that can be used to improve the usability of a product. We reasoned that the basis for product improvement is through changes that are linked to recommendations that are linked to the events that occurred during the formative usability test. Furthermore, we realized that, as with other elements of a report, the context of use of the report will have a strong influence on what metrics to report and how to report them. For example, practitioners will provide the metrics that they are paid to

REPORTING ON FORMATIVE USABILITY TESTING

provide. In the case of consultants, this may be only the high-level metrics, which might only appear in the rationale associated with recommendations. For in-house practitioners, it appears that it is a common practice to include measurement detail in appendices for archival purposes.

The model below shows how event-level details feed into recommendations and their rationales. Though recommendations are shown at the top of the figure, we did discuss the possibility that a practitioner might not be in a position to offer good recommendations (for example, if the practitioner is new to the technical field). If practitioners can offer detailed recommendations for change, then they should unless there are overriding considerations to avoid doing so (for example, a consultant whose contract is only to provide descriptions of events). We didn't have time to discuss this in more detail, but it seems that it should always be possible to provide a recommendation, even if the recommendation is "Product developers should investigate and provide recommendations (or implement changes) to eliminate or reduce the impact of this event."



Ideally, the recommendation should describe one or more suggestions about how to fix the problem, and the rationale should include some description of what happened and why it happened (so one way to quickly express the scheme is: WHAT + WHY = HOW).

Performance on the Expected Path refers to measurements made when there is an expected path – for example, (1) did the event lead to task failure, (2) if not failure, was it difficult for the participant to recover, or (3) was recovery fairly easy? The position paper distributed before the meeting has preliminary definitions for these various types of behavioral data. The point was raised during discussion that practitioner judgment also plays a role in the rationale for recommendations (for example, knowledge of standards, experience about elements of design that might not have a devastating effect on user performance but are known to lead to significant criticism in product reviews, etc.).

There was general agreement that practitioners should report the core information – the recommendation and its rationale, citing supporting evidence in the rationale. If

appropriate, practitioners can report a set of recommendations to address a particular event or class of events, ideally, organized in a prioritized order.

Rules Breakout Group

How we worked on the rules

Prior to the workshop, each of the common elements as defined in the October workshop was printed on a purple post-it note. Each of the 88 elements was printed on a purple post-it note. All of the rules (165) formulated in the October workshop were printed onto yellow post-it notes. The 88 purple post-its representing elements were posted to the wall. Each workshop participant in the rules breakout group was given a set of the labeled yellow post-its representing rules to associate with a common element.

When this exercise was completed, the group examined those elements that did not have any associated rules and developed rules indicating when to include this element. Participants were asked to review the elements and rules and generate any new rules that were missing. Finally, participants reviewed the rules to distinguish and label those that described how to use the corresponding element(s), rather than when. Those that described how to use the element(s) were identified with a blue dot.

Once this phase was completed, participants held a group review and discussion of the rules to clarify or expand on individual rules. For instance the rule: "If no one's going to read it, don't write a report" stimulated a discussion addressing the question "Is there ever an instance in which a report is not written?". Participants discussed what constituted a report. Is an informal email that lists the top 10 issues to be addressed by developers a report? Is a debriefing or presentation after a test a report?

Rules for general characteristics of the formative report

1. The first set of rules discussed was under the heading "General" referring to general characteristics of the formative report. The rules listed below were identified for this category: If the audience for the report is unknown; assume a broader audience and include more types of information.
2. If multiple audiences need to be addressed; segment elements for different readers.
3. If the organization's culture has certain expectations; identify them, and try to meet those expectations.
4. If you don't know your audience; be formal: include a lot but use plain language.
5. If the report is oral/PowerPoint; create it for stand alone use.
6. If no one will read it; don't write a report.

REPORTING ON FORMATIVE USABILITY TESTING

7. If there is a lot of trust of the design/test team; keep it simple.
8. No one may read it now, but they may in the future
9. If the audience is not familiar with usability/experimental human factors jargon; try not to use overly academic sounding terms and define terms.
10. If you are doing frequent, iterative test and report; be consistent in the reporting components you use.
11. If someone else can capture what happened; let them write the report.
12. If you are delivering to managers or developers, make it as short as possible.
13. If no one will read it; write a shorter report.

Rules for positive findings

The next report element examined by the group was “Positive Findings”. Listed below are the rules associated with this element are:

1. Persuasion is a goal; focus even more than usual on positive findings.
2. Need to not turn developers off or need to give them encouragement; start with positive results.
3. Always; include positive findings.
4. Unless the audience is very thick skinned; always include the positive findings.
5. The things you hoped would work actually worked; include positive findings.
6. Really bad news; includes positives.
7. If the news is really bad; focus on “hope” recommendations.

Rules for detailed findings and problems

The final element discussed was “Detailed Problems/Findings”. The rules identified for this element include:

1. If the training team is using the document; indicate what “features”/problems should be included in training.
2. If the help team is using the document, give a list of user synonyms for all items or tasks for inclusion in the index.

REPORTING ON FORMATIVE USABILITY TESTING

3. If the help team is using the document; indicate what problems need to be addressed in the help.
4. If user articulates any assumptions; call out those assumptions and whether these are accurate. E.g. "I'm hoping it will save automatically".
5. Audience is upper management; more visual elements (charts, screen shots), fewer words.
6. You need to get your audience's attention; start with key "tough issue/problem" (and a possible recommendation or alternative to give hope).
7. If you are a newer usability person (or new to team); less directive language and recommendations.
8. If there is bug tracking software you may include the bug "numbers" in the report. (Also link to online report in the bug description.)
9. Know team is aware of problems; don't focus on them.
10. Credibility of findings is an issue/priority; align findings and interpretation and recommendations with lit/citations.
11. If purpose is archival and audience is UCS savvy; use format for research reports in literature.
12. Report minor issues/quick fixes; organize by page or screen.
13. Want to emphasize UI pieces that don't work; report on number of failures/task abandoned.
14. Audience is not "in touch" with users; include more subjective findings to build awareness (and empathy?).
15. Development is moving quickly; limit the number of findings reported.
16. If they needed help; note this.
17. Executives are the primary audience; define severity priority and other codings in comparison to business goals or product/project goals.
18. Issues and design impact and complex; consider interactive, non-linear presentation of findings and recommendations.
19. If engineers; use a tabular format
20. If there is a lot of distrust of the design/test team; emphasize findings build consensus on them before recommendations.
21. Team responsibilities are well know; organize findings based on who's responsible for them.
22. Using results for discussion; include problem descriptions.

REPORTING ON FORMATIVE USABILITY TESTING

23. Don't confuse participant comment with "problem" that has behavioral consequences.
24. You were testing to answer specific questions; report findings relevant to these questions.
25. Always; distinguish from aspect of UI that caused behavioral consequences and those that did not (e.g. std violation misspellings, etc.)
26. For a persuasive report; do not follow CIF format – any info not needed to understand findings (user profiles, method) should be in a appendices.
27. Prioritizing happens as team activity; report these priorities. Don't augment or reprioritize.

4 Next Steps

Common Elements and Rules

We would like to develop a set of guidelines to give to practitioners about reporting formative testing.

Based on the work in both the October 2004 and this workshop, we believe that the elements and the rules will be best presented if they are connected to good examples from the collected sample reports.

The next step is to design a way to present this information effectively. We will start by taking the most “solid” subset of the elements and map them to the examples. This subset will be used as a prototype of how to present the information

We expect that this work will help us determine the next steps on how to proceed with the development of guidelines.

Reporting Context

Reporting context continues to be difficult to pin down. The interviews done during this workshop still need to be reviewed. We hope to use information and anecdotes from them to provide additional data on the contextual issues that are the most important in planning a report (or series of reports).

Metrics Group

The next step will be to write systematic rules (guidelines) for reporting metrics, similar to those being written for other aspects of formative usability test reporting. There will be some significant challenges due to variability in practice among experienced practitioners. We need to take care to accommodate all reasonable variation. Recommendations for reporting metrics depend on having some idea about what practitioners choose to collect as data and how they choose to analyze them.

It seems reasonable to start with high-level decisions about what types of measurements to include or exclude in a report. Once we have an idea about when to include a type of measurement, we can move to guidelines for (1) mathematical treatment – e.g., computation of means or medians or both, (2) statistical treatment – e.g., reporting measures of variability such as the standard deviation or computing confidence intervals, (3) analytical techniques (e.g., verbal protocol analysis) to help forge the link between data and recommendations, expressed in the rationale for recommendations, and (4) reporting techniques – e.g., in text, in tables, in graphs, or some combination (and if in graphs, guidelines about the use of which types of graphs for which types of data).

There’s still a lot of work to do.

5 Appendix: Related Materials

Questionnaire on Context and Details of the Reports

Seven people completed questionnaires. Their responses are shown here.

How are you employed	<input checked="" type="checkbox"/> [6] Employee or place full time in a company, not primarily in usability field <input type="checkbox"/> [1] Employee of a usability or user experience company <input type="checkbox"/> [0] Consultant or contractor <input type="checkbox"/> [0] Academic researcher
What was the method or technique used in the evaluation this report covers	<input checked="" type="checkbox"/> [4] Formal lab usability test <input checked="" type="checkbox"/> [2] Informal usability test <input type="checkbox"/> [0] Card sorting <input type="checkbox"/> [1] Other
What material was used for this test?	<input type="checkbox"/> [1] Fully functioning product <input checked="" type="checkbox"/> [2] Partially functioning product <input checked="" type="checkbox"/> [4] Prototype: on the computer <input type="checkbox"/> [0] Prototype: paper
Who was the audience for this report (Check all that apply)	<input checked="" type="checkbox"/> [6] Management or other executives <input checked="" type="checkbox"/> [4] Development or product team <input checked="" type="checkbox"/> [5] Design or usability team <input checked="" type="checkbox"/> [2] This was a formal record of the test
What is the style of the report	<input checked="" type="checkbox"/> [6] Document <input checked="" type="checkbox"/> [2] Presentation <input type="checkbox"/> [1] Video
How large was the team that worked on the evaluation this report covers Do not include people who simply observed the test	<input checked="" type="checkbox"/> [2] 1 person <input checked="" type="checkbox"/> [3] 2-3 people <input checked="" type="checkbox"/> [2] 4-5 people <input type="checkbox"/> [0] More than 5 people
Did people from the development team, business unit or management observe the evaluation?	<input checked="" type="checkbox"/> [6] Yes <input type="checkbox"/> [1] No
How long after the evaluation was this report completed	<input checked="" type="checkbox"/> [2] Within a week <input checked="" type="checkbox"/> [4] Within two weeks <input type="checkbox"/> [1] Within a month <input type="checkbox"/> [0] Over a month later
Is this report part of a formal development or usability methodology in use at the company?	<input checked="" type="checkbox"/> [2] Formal software development methodology <input checked="" type="checkbox"/> [3] Formal usability methodology
Is this report based on a template?	<input type="checkbox"/> [0] Template, used for all reports <input type="checkbox"/> [1] Template, but modified as needed <input checked="" type="checkbox"/> [3] Based on a general outline, but not a template <input checked="" type="checkbox"/> [3] No

Common Elements in Formative Reports

Consolidated list of elements, based on David Dayton's analysis, June 2005

Title page and global report elements

E1	Title page or area
E2	Report author
E3	Testers names
E4	Date of test
E5	Date of report
E6	Artifact or product name (with version ID)
E7	Table of contents
E8	Disclaimer
E9	Copyright/confidentiality
E10	Acknowledgements
E11	Global page header or footer

Executive Summary

E12	Executive Summary
-----	-------------------

Teaching Usability

E13	How to read the report
E14	Teaching of usability

Introduction

E15	Scope of Report
E16	Description of artifact or products
E17	Business goals and project
E18	Test goals
E19	Assumptions
E20	Prior test report summary
E21	Citations of prior/market research

Method and Methodology

E22	Test procedure or methods used
E23	Test protocol (including prompts and probes)
E24	Scripts
E25	Limitations of study or exclusions
E26	Data analysis description

Overall Test Environment

E27	Overall Test Environment (Lab, field, remote, etc)
E28	Software test environment (OS, Browser, etc)
E29	Screen resolution or other critical description
E30	Data collection and reliability controls
E31	Testers and roles

Participants

E32	List or summary of participants
E33	Number of participants
E34	Demographics or specific background
E35	Relevance to scenarios
E36	Experience with product
E37	Company, employee (years in current job)
E38	Educational background

REPORTING ON FORMATIVE USABILITY TESTING

E39	Description of work
E40	Screeners

Tasks and Scenarios

E41	Tasks
E42	User-articulated tasks
E43	Scenarios
E44	Success criteria
E45	Difficulty
E46	Anticipated paths
E47	Persons on the task

Results and Recommendations

E48	Summary
E49	Positive findings
E50	Observed issues and recommendations (table)
E51	Problems / Findings
E52	Recommendations
E53	Definitions of coding schemes

Detail of Recommendations

E54	Severity of errors
E55	Priority
E56	Level of confidence
E57	Global vs specific
E58	Classification as objective and subjective
E59	Reference to previous tests (better or worse)
E60	Bug or reference number
E61	Fix - description and effort

Metrics

E62	Performance data / summary (pending metrics)
E63	Satisfaction questionnaire results (pre or post-task or test)

Quotes, Screenshots and Video

E64	Quotes
E65	Screenshots with callouts
E66	Video clips
E67	Voice-over

Conclusions

E68	Discussion
E69	Interpretations or implications for use of findings
E70	Tie-back to test or business goals
E71	Lessons learned (test process, product)

Next Steps

E72	Further studies recommended
E73	List of what needs to be done
E74	Ownership of issues
E75	New requirements and enhancements
E76	Style guide updates

Appendices

E77	Test materials
E78	Detailed test protocol or scripts

REPORTING ON FORMATIVE USABILITY TESTING

E79	Tasks and/or Scenarios
E80	Consent Forms
E81	NDA
E82	Preliminary Report
E83	Data analysis
E84	Raw data
E85	Data repository
E86	Data retention policy
E87	Statistics
E88	Style guide (for product)

Proposed Rules for Reporting Formative Usability Results

Classification of rules suggested at the October workshop in Boston
 Analysis, Whitney Quesenbery - June 10, 2005

Initial Decisions: Planning the Report for Your Audience

Deciding on a report format and formality

R1	Always	Identify the most important audiences who will read the report
R2	Audience for the report is unknown	Assume a broader audience and include more types of information
R3	If multiple audiences need to be addressed	Segment elements for different readers
R4	Organization's culture has certain expectations about deliverables	Identify them, and try to meet those expectations
R5	The goal is to move quickly to next iteration and there is buy-in to do this	Don't need a formal "report". Instead just list issues to address
R6	The goal is to rapidly iterate the design (and this is already in the project plan)	Keep the "report" informal / as short as possible and focus effort on design instead
R7	You don't know your audience	Be formal: include a lot but use plain language.
R8	If turnaround is critical (no time)	Restrict report to exec summary and findings table
R9	If the report is oral PowerPoint	Create it for stand alone use
R10	If report is part of other communication products	Reference (link) to these other documents
R11	If no one will read it	Don't write a report

Reporting to executives and other managers

R12	Executives are the primary audience	Define severity, priority and other codings in comparison to business goals and /or product/project goals
R13	Audience is upper management	More visual elements (charts, screen shots), fewer words
R14	If need to persuade a committee or stakeholder group or busy exec	Present key findings in a presentation not a document

REPORTING ON FORMATIVE USABILITY TESTING

R15	If the audience has limited time (decisions makers)	Make the critical elements of the report prominent and easy to read
R16	Development managers are primary audience	Define codings (severity, priority, risk, errors) in terms as comparison to product/project goals
R17	The decision maker is not actively involved in the testing	Include more detail and persuasive arguments
R18	Education of decision makers about importance of user involvement	Write results to show the user input directly (e.g. with quotes or before/afters)

Adjusting the report if the analysis is done by the whole team

R19	The team is involved in collecting observations or debriefing	Report on names of team members who participated in watching or debriefing
R20	Results are team driven	Summarise arguments
R21	Diagnostics carried out by team members	Focus on action-related communication
R22	Prioritising happens as a team activity	Report these priorities. Don't augment or reprioritise

Considering how this test fits into the overall UCD/development cycle

R23	It is early in development cycle	Focus on concepts not metrics
R24	Test occurs shortly before planned rollout	Include both quick fixes and longer-term recommendations
R25	Prototype changes between tests	Note which version observations pertain to
R26	Frequent, iterative test + report	Be consistent in the reporting components you use
R27	Development is moving quickly	Limit the number of findings reported

Managing Politics: Ensuring Acceptance of the Report

Getting the audience in a receptive frame of mind

R28	You need to get your audience's attention	Start with key "tough issue/problem" (and a possible recommendation/alternative to give hope)
R29	Persuasion is a goal	Focus even more than usual on positive findings

REPORTING ON FORMATIVE USABILITY TESTING

R30	Need to not turn developers off or need to give them encouragement	Start with positive results
R31	Always	Include positive findings
R32	Unless the audience is very thick skinned	Always include the positive findings
R33	The things you hoped would work actually worked	Include positive findings
R34	If there is a lot of trust of the design/ test team	Keep it simple
R35	Really bad news	Include positives
R36	If the news is really bad	Focus on “hope” recommendations

Establishing credibility

R37	Need to impress audience with your knowledge of the field	Use jargon (heuristics). Cite references.
R38	Need to impress audience with your knowledge of the field	Consider what bias you are adding to the findings
R39	If there is a lot of distrust of the design/test team	Emphasise findings, build consensus on them before recommendations
R40	Credibility is an issue	Refer to, emphasize performance metrics (objective measures) if possible
R41	Credibility of findings is an issue/priority	Align findings and interpretation and recommendations with lit/citations
R42	You need to explain the rationale for design decisions after an iterative process	Create a representation of each iteration and key outcomes of each test
R43	If you are a newer usability person (or new to team)	Less directive language and recommendations
R44	You don't know all your audience well.	Declare your biases and test limitations

Communicating strategy and big picture

R45		Give the big picture (for example the tested product is good, the following list of recommendations will produce incremental improvements)
-----	--	--------------------------------------------------------------------------------------------------------------------------------------------

REPORTING ON FORMATIVE USABILITY TESTING

R46	Major issues are found suggesting a need for a change strategic direction	Focus on high level issues and recommended actions, not screen-by-screen or minor issues
R47	If you give fundamental design/ Strategic recommendations that cannot be implemented in this release	Include this information with roadmap for next release
R48	Issues and design impact are complex (e.g. multiple problems / changes that impact flow, architecture, etc.)	Consider interactive, non-linear presentation of findings and recommendations

Using the report to move the work to the next step

R49	You want to prompt discussion	Include questions (preferably as headings or in a discussion section)
R50	Using results for discussion	Include problem descriptions
R51	Issues become “to do” items for various members	Format them as a spreadsheet or checklist that is easily sorted
R52	?	Include “Action-List” that can be used to (a) Track actions taken (b) assign responsibility / accountability for usability follow – through
R53	Team responsibilities are well known	Organise findings based on who’s responsible for them
R54	Need to record next actions/owners	Use a table format?
R55	During discussion, team proposes subsequent testing	Record UI elements or proposed task to test
R56	Get new ideas during test	Call out new ideas in report
R57	If it is important to “sell” more testing	Include “next steps” persuasion/sales
R58	If you collected feedback for other product teams	Collect those in one section of the report

Planning for future use of the report (and archiving)

R59	You need to go back and revisit data in future	Formalise/organize data and findings notes repository
R60	If the report has shelf life	Include background methodology etc

REPORTING ON FORMATIVE USABILITY TESTING

R61	Purpose is archival	Provide enough methodology detail for someone to duplicate the study
R62	If purpose is archival and audience is UCD savvy	Use format for research reports in literature
R63	No one will read it now but in the future they might	

Structuring the Report: Assembling the Contents

Best practices and elements to always include

R64	Never	Tell me to always do something
R65	Always	Include a table of contents
R66	If report informal (less than X pages)	Do not include ToC
R67	Always	Include an executive summary that at least mentions all high priority recommendations
R68	Always	Include exec summary
R69	Always	Keep exec summary brief! 2-3 page exec summaries are not summaries.
R70	Always	Include test objectives/goals
R71	Always	State the purpose of the evaluation
R72	Always	Describe product or part being tested
R73	Always	Include the scenarios (can be appendix or in report) to set context for findings

Organizing principles

R74	For a persuasive report	Do not follow logical CIF format – any detailed information not needed to understand the findings (eg user profiles and method) should be in the appendices
R75	If the report has to be easy to read	Put the results in the content and the rationale and justification in appendix

Describing the scope or limits of the test?

R76	You were testing to answer specific questions	Report findings relevant to these questions
-----	-----------------------------------------------	---------------------------------------------

REPORTING ON FORMATIVE USABILITY TESTING

R77	(You want to) emphasize parts of UI tested or not tested	Report on sections of software not used during test
R78	To emphasize parts of UI tested or not tested	Identify sections of software not used during test

Reporting on test goals and methods

R79	Disagreement about any aspect of methodology	Include details about the disputed part of the process
R80	If a known bias exists in method (users out of profile, etc.)	Include caveats with the data (or discard)
R81	Audience unfamiliar with method (e.g. paper prototyping)	Include brief overview
R82	The audience is unfamiliar with usability testing	Include more description of the method and what usability testing is
R83	Audience is not familiar with usability / experimental human factors jargon	(a) try not to use overly academic sounding terms (b) Define terms
R84	If "management" needs to know	describe process
R85	If first in series	Include methodology section
R86	If one in series of tests	Be brief don't repeat methodology
R87	Team observed tests	don't include methodology
R88	If process or methodology is already known	Delete it

Describing user tasks

R89	If application is designed for work	Focus evaluation on work tasks, test in context of work tasks
R90	If user supplied the task	report what they said they would do
R91	Disagreement about the tasks	Include description of tasks

Reporting information about participants

R92	If multiple user groups	Include demographic survey or results that distinguish them
R93	If a small subset of the user population is recruited	Always include particulars about each User (not just user profile)

REPORTING ON FORMATIVE USABILITY TESTING

R94	If you have access only to internal users	Use them with great caution! (And acknowledge risk)
R95	Test participants include special needs/disabled users	Don't always segment isolate them in findings
R96	Testing internal users (e.g. surrogate users)	Note that these were internal users
R97	Significant differences in users tested	Describe users tested

Reporting problems and observations

R98	want to emphasize UI pieces that don't work	report on number of failures/task abandoned
R99	If product failed or there were known bugs that affected the test	note the problems and which observations may be in doubt
R100	always	Distinguish from aspect of UI that caused behavioural consequences and those that did not (e.g. std violations, misspellings, etc.)
R101	If they needed help	Note this
R102	If they used doc	Note this
R103	?	Don't confuse participant comment with "problem" that has behavioural consequences
R104	If user articulates any assumptions	Call out those assumptions and whether these are accurate
R105	?	Should the length of the report be limited by excluding some low priority observations?
R106	Outliers	How to decide whether to include problems only encountered by one person?
R107	If ?	Should all observations be included or should the length of the report be limited?
R108	Team accepts usability specialists ability to categorize	Categorize the results
R109	If observations are categorized or grouped	Include explanation of how categories were derived
R110	Reporting minor issues/quick fixes	Organize by page or screen
R111	Want to get team buy in	Use categories supplied by team

Creating and Reporting Recommendations

R112	Only development team is using findings	Build recommendations through consensus and design activities
R113	If?	Organize the recommendations into categories related to the product architecture
R114	If providing recommendations	describe any caveats or situations that might negate them
R115	If report includes recommendations	Note whose recommendations they are (usability person, team, individual developer)
R116	If recommendations are required	Link recommendations directly to observations when possible
R117	You are not familiar with the domain or technology	Qualify recommendations or provide alternatives (“If X isn’t possible, consider Y”)
R118	If architecture is complex / you aren’t decision maker	couch recommendations as suggestions
R119	Always	Give an identifier to each recommendation
R120	Test participants include disabled/special needs users	Consider recommendations that highlight those needs, and whether compromises result
R121	?	Don’t include recommendations unless audience is expecting them
R122	Re-design will be addressed by someone else	Recommendations can be more vague and include things like “consider ...”
R123	If development process assumes you provide design recommendations	Give them Else focus on describing problems
R124	Always	Include recommended solutions for each problem

Reporting on Metrics: Task Success, Severity Ratings, Errors

General considerations

R125	The performance metrics are split or close or not clear	Use more conditioned or softer language if you include recommendations
R126	The performance metrics are split or close or not clear	Focus on goal alignment and interpretation, rather than metrics details

REPORTING ON FORMATIVE USABILITY TESTING

R127	Team gets a little too focused on metrics and stats	Don't include stats
R128	If metrics are used	Present them in visual form : table, graphs, spreadsheets
R129	Your time is very limited	Omit counts (or track them in some automated manner)
R130	Always	Include a frequency count for observation (9 of 10)
R131	Fewer than X participants	Don't report percentages e.g. 3 out of 5 ≠ 60%
R132	If the test is with small number of users	Count number of users experiencing each problem
R133	Credibility building is important	Increase metrics or "analytical" representations (charts, tables)

Reporting task success rates

R134	if assisted participant	No task success
R135	If assisted participant and they complete task	Mark task as success with assist

Reporting "time on task"

R136	If not running system	Don't use time on task
R137	If verbal protocol used	Do not include time on tasks
R138	If task failure	Exclude from mean task success time

Counting errors

R139	If insufficient sample size	Don't report number of errors and number of times error made
R140	Always	report number of errors per task
R141	Always	Report number of times error was made

Using severity ratings

R142	Bug tracking software is used and usability issues can be tracked in it	Use a format and priority/severity ratings that will fit in with that existing infrastructure
R143	If recommendations are required	give them severity ratings
R144	You use coding (severity, priority, risk, error level, value, etc.)	Get consensus or definitions before using

REPORTING ON FORMATIVE USABILITY TESTING

R145	If you are giving severity information	Give the big picture stuff. How good is this product relative to something
R146	If severity ratings are used	High = serious obstacle to use, Low = cosmetic or nice to have and quick fix
R147	Always	Supply severity ratings for problems and priority for fixes
R148	If reporting severity	Explain severity scale
R149	Audience capable of assessing severity	Omit severity ratings
R150	Audience has sufficient background to make informed decisions	Omit severity, prioritizing

Presentation: General issues

Including highlights video, transcripts and verbatim quotes

R151	If you have time	Add a highlight video
R152	If possible and useful	Include quotations from users associated with each observation
R153	Want to be persuasive	Include quotes
R154	Is there ever a reason to	Include transcripts
R155	Persuasion is a goal	Include user quotes, video clips, etc
R156	If you have time	Consider adding highlight videos
R157	Persuasion is a goal	Include video clips
R158	Audience is not "in touch" with users	Include more subjective findings to build awareness (and empathy?)

Deciding to use screenshots

R159	Audience all familiar with interface	Screenshots optional
R160	If readers don't know interface	include lots of screenshots
R161	Development team is (folks who will implement change) remote from usability person	Include screenshots or clips
R162	To illustrate cumbersome process	Use sequence of annotated screen shots
R163	Top issues are strategic or conceptual	Don't include screenshots if they will distract from the main issues

REPORTING ON FORMATIVE USABILITY TESTING

R164 To increase readability Illustrate all observations/recommendations with a screen shot

Writing for good readability

R165 Easy to read report Lots of bulleted lists – no long paragraphs

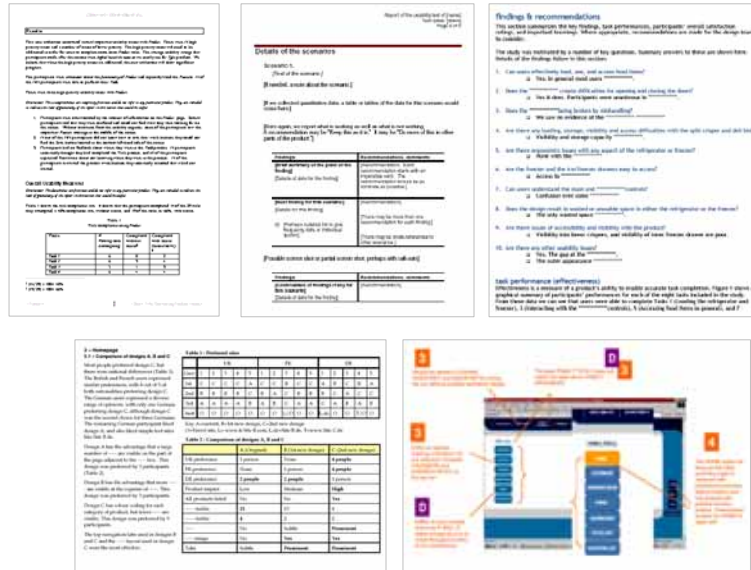
UPA 2005 Workshop on
Reporting Formative Usability

Elements from the Sample Reports and Templates

Adapted from presentations at
Boston Workshop, Oct 18-19, 2004

Whitney Quesenbery
Ginny Redish

A range of styles



Executive summaries

Came in all sizes, layouts, approach and details, but were always first

executive summary
 This formative usability evaluation examined users' behaviors when asked to perform a series of tasks with a prototype version of the refrigerator. Ten refrigerator owners representing the market segment participated in the study.

The most positive user feedback was for the perceived space, visibility, accessibility and brightness of the refrigeration cavity. The lack of ice/water dispenser was a major dissatisfier for most. Some specific usability problems were identified and the top few are shown here:

- Closing the refrigerator doors
- Risk of items falling behind freezer drawer**
- Mapping ingredient controls to bins
- "Invisibility" of internal Freezer drawer
- Lower crisper bin capacity
- Confusion over
- Unable to find water filter
- Some button labeling

Executive Summary

Problem was tested to evaluate *Features* that were new to this build. Previous testing of this product had indicated problems with *Features* that were redesigned and also addressed in this evaluation. NN participants were obtained from a recruitment agency, who had N years background in *Appliances* *Domain* and N years *Company* experience. Participants performed n tasks covering *work process 1*, *work process 2*, and *work process 3*, which were available in this version of *Product*. The results indicate three main usability problems: *problem 1*, *problem 2*, *problem 3*. Participants were able to complete tasks without assistance about 75% the time. They rated the product *above/below/neutral* on usefulness and attractiveness, and *above/below/neutral* on ease of use and clarity. A total of NN usability issues with design recommendations are described. Future usability testing when the product is near release and another user evaluation for the following point release are planned.

Executive Summary

A usability study of the *smokefree.gov* site was conducted by the Communication Technologies Branch of the National Cancer Institute on March 18th, 2003. Five participants, with various skill levels and web experience completed usability testing of the website in five separate sessions. Detailed participant profile information is presented in the 'Methodology' section, under 'Participant Demographics.'

Each participant completed a set of typical tasks on *smokefree.gov*, thinking aloud as they worked. Task list, including preferred path and answer, is in 'Appendix B: Task List.' Performance and subjective results were recorded by the usability team while observing participant interactions.

Most significant usability findings:

- From the home page, participants did not know what sort of information to expect in the step-by-step guide.
- Participants could not find information easily in the step-by-step guide.
- Participants did not immediately notice or use the left navigation bar. In the online guide.
- Participants did not use the tabs across the top of the pages. They did not use it to navigate between pages nor did they use it to get back to the home page (they simply used the back button to get back to the home page).

The complete set of evidenced-based usability findings is presented in 'Findings: Detailed Usability Findings.'

After completing the tasks, participants shared opinions of their experience with the site using verbal debriefing. They also completed a post-test paper questionnaire. A summary of the subjective feedback is presented in 'Findings: Participant Preferences Summary.' Detailed participant responses to the post-test questionnaire are in 'Appendix C: Post-test Questionnaire Responses.'

Key findings summaries

Summary of Usability Findings

The [www.XXXXX](#) usability evaluation resulted in 40 major findings and corresponding research-based recommendations. Each finding was given a severity rating of Low, Medium or High. Following is a breakdown of the findings by severity:

- **Low:** 10 Findings
- **Medium:** 12 Findings
- **High:** 18 Findings

During the usability evaluation, 38 tasks were attempted. Only 19 of these tasks were completed successfully. This calculates into a 50% overall success rate for the site. In addition to the task completion rates, satisfaction rates were calculated from the post-test questionnaire. The overall satisfaction rating for the site was 59%.

Please see the Findings section of this report for detailed task analysis and satisfaction ratings.

Key findings from 4/18 usability test

Home Page!

- Home page is still busy. People don't see a distinction between the contents of the left and right columns.
- People are scanning the home page for something that reflects their current situation or issue (not self-identifying). Sometimes this causes them to choose the wrong thing.
- The sections for people not receiving benefits were not clearly identified as such. Specific issues included:
 - The link for "learn about" was unclear, or overlooked
 - Where would you go if you are in the application process or have been denied?
 - The link for "Plan your financial future" did not clearly indicate that it was for people not receiving benefits (the intended users of that link)
- Many people have questions related to their spouse, and these issues are not addressed clearly at the site (e.g. spouse's benefits, death of a spouse, divorce)
- Things not found on the home page or easily at the site:
 - Things you can do online
 - Medicare information
 - Local offices

Overview of success rates for website

- Overall success rate across all the people and tasks: **81%**
- Updating **membership** information was easy: 100%
- Finding **people** - members and **xxxx** staff was easy: 93-98%*
- Finding out how to **join** was relatively easy: 88%
- Finding the contents of **past meetings** was harder: 63-68%*
- Getting to the part of the site to make a personal meeting **schedule** was easy: 96%
- Actually developing the schedule was harder: 68-93%*
- Finding the latest **news** was easy: 92-93%*
- Finding print **publications** and related resources was harder: 47-81%*

5

Only problems? or positives, too?

Many include both positives and problems/issues.

positive findings

Participants were unanimous in their praise and delight at the *****. The following were all well received:

- Overall exterior design
- Capacity and visibility of RC
- Height of RC
- Split crisper bins
- Easy to pull out FC
- Easy to understand RC/FC temperature controls
- Easy to understand Water Filter status
- Easy to clean shelves
- Touch buttons and appearance of controls

Reaction to the new design

Main results

- All evaluators recognized it as a wizard without being told.
- All evaluators could use the navigation bar
- The next step link is still not working. This must be made visually stronger
- Easier to implement with the one screen design
- The design will be easy to make generic

Some open issues remain to be refined

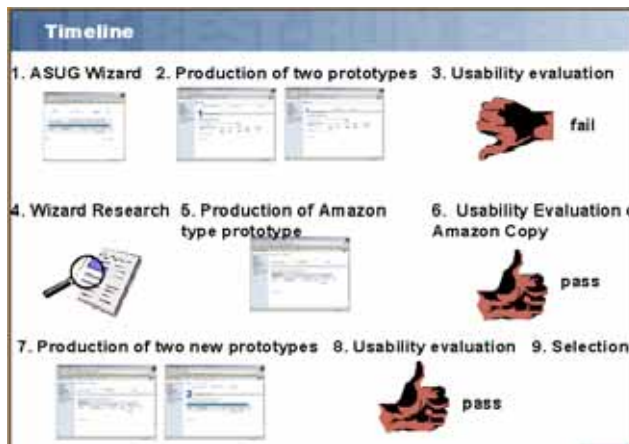
The design functions but there are some minor open issues. These and other issues will be discussed after testing in a refinement meeting (Jan 03). Examples of open issues:

- Rather than (0) for shopping cart change it to (empty). 0 is too technical if we were to look at Amazon it would say empty.
- Color of the post number. I believe it should be an orange but further testing will verify this finding.
- The text describing each stage should not overhang the wizard end lines. At the moment the first step text overhangs.

	CSH
Successes:	Noticed the tool tip on compensator gain
	Used help button in dialog box to get information
	Knew to turn WTC mode on and off
	Seemed to move from CSH to HB easily and without confusion
	Followed link to example from root-locus CSH. Understood what would
	Read through all the information on the CSH screens (wasn't concerned
	length)
	Used CSH on root locus diagram to get definition. Followed link, actually
	the doc
Issues:	CSH should link to example that uses tool, not one that doesn't
	See also should link to closely related topics
	Looking for info/overview on the different systems - wants it from dialog

6

Explaining the UCD Process



This presentation reported on a series of usability evaluations

7

Goals for the test

Some stated clear goals for the specific test, others had only general usability goals or no explicitly stated goals

Purpose of Usability Evaluation

The purpose of the [www.XXXXX](#) usability evaluation was to:

- measure users' abilities to perform tasks on the site, (*performance evaluation*)
- obtain users' impressions of the website, (*preference evaluation*)
- identify difficulties involved in using the site, and
- suggest research-based recommendations for improvement.

The intent is to ensure that [www.XXXXX](#) is:

- easy to learn and use,
- useful to the intended audience, and
- satisfying to use.

The findings and recommendations of this evaluation are intended to be able to effectively and efficiently satisfy their goals.

research objectives

research questions

The following lists some specific research questions that the study will address:

1. Can users effectively *****?
2. Does the *****create difficulties for opening and closing the doors?
3. Does the *****risk being broken by mishandling?
4. Are there any loading, storage, and access difficulties with the *****?
5. Can users easily reach and understand all of the main and secondary controls?
6. Are there ergonomics issues with any aspect of the *****?
7. Is the ***** easy to access?
8. Can users understand the main and *****controls?
9. Does the design result in wasted or unusable space in either the *****?

8

Participant demographics

Participant Demographics

Participants having the following profile characteristics evaluated Smokefree.gov

#	Age	Smoker	Last attempt to quit smoking	Race	Education
1	37	Yes	Less than 6 months	Caucasian	Some college or college degree
2	39	Yes	6 months-1 year	Caucasian	High School graduate/GED
3	67	No (Not a smoker)	-	Caucasian	Some college or college degree
4	30	No (Not a smoker)	-	Asian	Some college or college degree
5	44	Yes	Less than 6 months	Hispanic	Some college or college degree
6	62	Yes	Less than 6 months	African American	Some graduate school or more

* Participant 5 did not know how to use a mouse and was not at all familiar with the web. So, this session was cut short and the data from the session was discarded.

User	Gender	Age	Ethnicity	Income	Ed.	Computer	Home type	Yard	Environ.
1	Female	55-64	Caucasian	40-50	BA	Basic	Suburban	Yes	Medium
2	Female	25-34	Hispanic	<40	BA	Medium	Apartment	No	Low
3	Male	65+	Caucasian	Retired	MS	Basic	House	Yes	Low
4	Male	35-44	Non-White	60+	BA	High	House	Yes	Low
5	Female	35-44	Caucasian	40-50	BA	Medium	Suburban	Yes	Medium
6	Female	55-64	Russian	60+	MS	Medium	House	Yes	Low

Table 1: Participant Characteristics

Participants
Total: 17

Gender	Count
Men	6
Women	11

Age	Count
under 25	1
25-34	2
35-44	9
50+	5

Race	Count
Asian	2
Black	5
White	9
Other	1

Education level	Count
Some high school	1
High school	1
Some college	1
College graduates	14

Note: One person spoke English as a second language.

Map of card sort participant characteristics. Text in yellow highlight is unchanged from persona description

Participant	Personas	Personas	Personas	Personas
Participant 1	Personas 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, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100	Personas 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, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100	Personas 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, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100	Personas 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, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100

Most used a simple chart.

One compared participants to personas being used for design

9

Scenarios or tasks

Listing of all task scenarios

Table 1: List of Task Scenarios

Task #	Description
1	<p>Loading the refrigerator and freezer:</p> <p>A) You have just arrived home from the supermarket with the bags seen on the table. You want to put all of the items away in the refr. Please go about loading the refrigerator just as you would normally home; feel free to sort the food items ... just whatever way you want.</p> <p>B) It is a few days later. You have just arrived home from the super bags of shopping you see on the table. You want to put all of the items in the refrigerator or freezer. Please go about loading the refrigerator just as you would normally do if you were at home.</p>
2	<p>Interacting with the main controls:</p> <p>You want to change the temperature of the freezer to make it as cold as possible. Please do this now.</p>
3	<p>Interacting with the storage bin controls:</p> <p>You want to change the settings on the storage bins so that storage is optimal. Please do this now.</p>

Full scenario details

Scenarios

To observe users addressing the issues identified as the goals for the test, the following scenarios were created

Scenario 1

You manage the network for TechMedia, a large software distributor.

To help you monitor the network, you've installed Software Development Company software on your PC in the Home office.

1. Locate the Software Development Company software on your desktop and launch the application.
2. Use the software to see what you can find on the TechMedia network. You have an IP range of 10.0.0.0. Among the devices you want to monitor will be FTP servers, web servers, mail servers, and other devices you choose to monitor. This may take several minutes.
3. When you have completed the new install wizard, take a few minutes to explore the application. Tell us what you see, what you think it does, things that you understand, and things that are unclear.

Information about you
e-mail address: marenan@SoftwareDevelopmentCompany.com
outgoing mail server: _____

Charts format

Appendix B: Task List

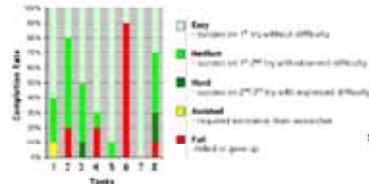
Task	Preferred Path and answer
1. You are considering quitting smoking and would like to know how to begin? What should you do first?	<input type="checkbox"/> Home <input type="checkbox"/> Online Quit Guide <input type="checkbox"/> Get Ready to Start <input type="checkbox"/> Answer: Set a quit date

10

...and with some visualization

Question	Agree	Neutral	Disagree
1. I think I would like to use this system frequently.	4	1	2
2. I found the system unnecessarily complex.	3	0	4
3. I thought the system was easy to use.	3	1	3
4. I think I would need support of a technical person to use this site.	2	1	4
5. I found the various functions in this system well integrated.	4	1	2
6. I thought there was too much inconsistency in the system.	4	1	2
7. I imagine most people would learn to use this system very quickly.	5	1	1
8. I found the system very cumbersome to use.	2	2	3
9. I feel very confident using the system.	4	0	3
10. I need to learn a lot of things to get going with this system.	1	2	4

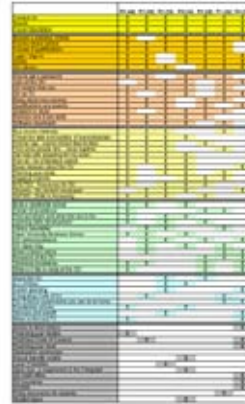
Task Effectiveness



Satisfaction



Appendix 3: Distribution of selected cards



13

Severity levels

Severity	Description
*****	A critical usability problem or user experience issue. User will be unable to advance to completion of task. Imperative to fix before release.
*****	A major usability or user experience problem. Product can still be used but user will have difficulty and might need to figure out a work-around. Fix should be given very high importance.
*****	An important usability or user experience problem. User likely to experience frustration but will be able to proceed. Cumulative effect will impact user perception of product. We strongly recommend that these be fixed.
*****	A moderate usability or user experience problem. Can be irritating to user and can negatively impact user's experience, but will not prevent task completion. Fix if possible.
*****	A minor usability problem or user experience problem. Cumulative effect of "ones" will damage perception of quality. Fix if possible.

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.	*****
4 - Problem described in Full Detail	Recommended usability solution described in full detail here.	*****

Recommendations	Priority
In view of the perception that Travel site is a ----- site, it is important to give users the expectation that fares are -----.	High
Make secondary navigation more prominent.	Medium
Signpost the range of content under -----, particularly singles holidays.	Medium
Provide an explanation of what ----- flights are.	Medium
A search box on the holiday homepage would simplify navigation.	Medium
Use tick boxes rather than radio buttons.	Medium
Clarify labelling.	Low
Signpost ----- if they are available.	Medium

14

Severity scales

When used, they were always defined

Severity Rating	How it is displayed	What it means
High	1 2 3	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> Participants were still able to accomplish tasks but had some difficulty doing so. Participants were frustrated or annoyed by the issue. Participants see their issue as a nuisance.
Low	1 2 3	<ul style="list-style-type: none"> Participants were able to use these features successfully with little to no difficulty.

- ***** An **important** usability or user experience problem. User likely to experience frustration but will be able to proceed. Cumulative effect will impact user perception of product. We strongly recommend that these be fixed.
- ***** A **moderate** usability or user experience problem. Can be irritating to user and can negatively impact user's experience, but will not prevent task completion. Fix if possible.
- ***** A **minor** usability problem or user experience problem. Cumulative effect of "ones" will damage perception of quality. Fix if possible.

usability severity ratings

- 5** A **critical** usability problem. User will be unable to advance to completion of task. Imperative to fix before release.
- 4** A **major** usability problem. System can still be used but user will have difficulty and might need to figure out a work-around. Fix should be given very high importance.
- 3** An **important** usability problem. User likely to experience frustration but will be able to proceed. Cumulative effect will impact user perception of product. We strongly recommend that these are fixed.
- 2** A **moderate** usability problem. Can be irritating to user and can negatively impact user's experience, but will not prevent task completion.
- 1** A **minor** usability problem. Cumulative effect of "ones" will damage perception of quality.
- D** A **graphic design/business** problem. This may adversely impact perception of quality and acceptance of the product.



15

Presentation formats for observations

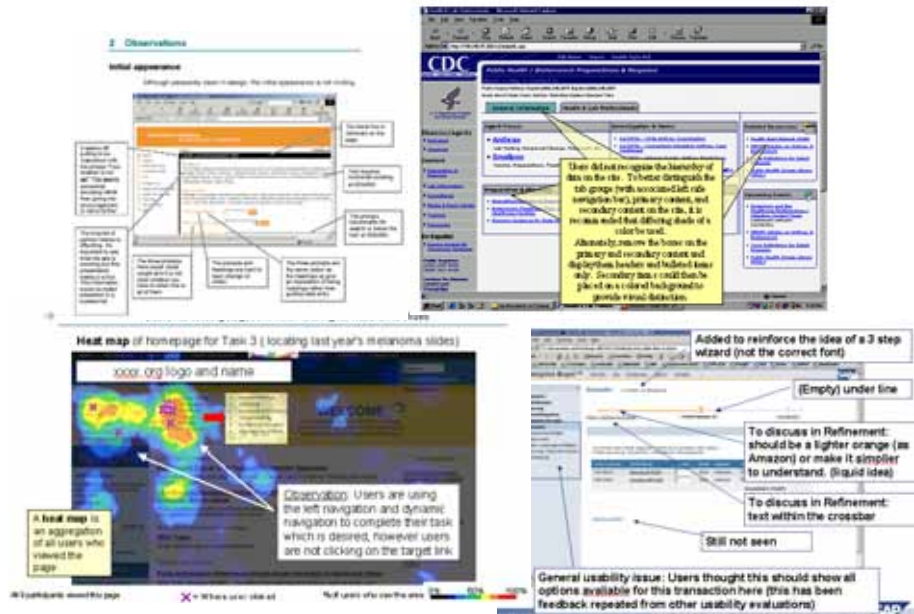
<p>Users did not notice or utilize self-navigation consistently.</p> <p>They only noticed it when they started searching for information and did not know where else to go.</p>	<p>Review the list of options, labels, words, a table of contents from which users will be able to be guided to find what they want.</p> <p>Each section should have the second level header as a prior cue to the top of that section/page.</p>
<p>Some participants were confused by the term "Register".</p>	<p>Use "Sign Up" instead of other words that users will understand.</p>
<p>Users ignored the option to view the right of some of the pages.</p>	<p>Review the path for technology that has to work. In general, use to the next language. Also, be sure to be consistent with the language throughout.</p> <p>Review the entire form. Incorporate content from the form into the page where it naturally fits in.</p>

"Starting Service" recommendations

- To improve access to starting services, provide a large, high contrast link or button to "Start/Stop Service". **Rebut:**
- To improve access to starting services, provide content that explains exactly what needs to be done. In addition to use cases #1 and #5 (Appendix B) users wanted:
 - easy instructions
 - what is picked up
 - car seats and options
 - pricing
 - specific times
 - phone numbers
- To improve access to related functions, consider separating the form into two separate forms that allow users to manage their accounts and report problems. **Rebut:**
- To improve access to related administration, provide links to "starting service" from the "Rates" and "Pay Bill" pages.
- To reduce confusion about where they are, strengthen the locational feedback of the top items (e.g. header content) and the side items (menu feedback than ensuring consistency).

16

Presentation formats for observations



17

Recommendations, separately

6. Recommendations

The following recommendations are organized by priority in terms of their usability implication. The meanings of the priorities are:

- 1 - Do not release the product without this change.
- 2 - Important to fix.
- 3 - Advisable if there is time and budget.
- 4 - Cosmetic changes.

Priority 1:

There should be some termination buttons in each screen of the setup wizard. These should be: Apply or Done, Cancel or Reset or Undo. The positive termination should bring up the next screen. Consider having the wizard steps or tabs at the bottom where the users attention is at the end of the dialog.

Selection of a template, in the parameters. The name and/or desc when the user works to modify the bring up the first setup screen of th

High Level Recommendations

- Provide a clear indication of the user's current location.
- Where possible, provide indications as to what the user should do next.
- Organize the content and functionality in a way that maps to an overall concept of the work.
- Use language that is commonly understood by the users.
- Use consistent language throughout the application.
- Keep the users informed about what is going on.
- Provide error messages in plain language, not codes.
- Provide error messages that describe precisely what and where the problem is and actions necessary to correct it.

When reported separately, recommendations were often by priority

18

Level of detail in reporting findings

Many give two level, with a summary followed by more detail, or organized in different ways

findings & recommendations

This section summarizes the key findings, task performance, participants' overall satisfaction ratings, and important learnings. Where appropriate, recommendations are made for the design team to consider.

The study was motivated by a number of key questions. Summary answers to these are shown here. Details of the findings follow in this section:

- Can users effectively load, view, and access food items?
 - Yes, in general most users.
- Does the ***** create difficulties for opening and closing the doors?
 - Yes it does. Participants were unanimous in *****.
- Does the ***** being broken by mishandling?
 - We saw no evidence of the *****.

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.	*****
4 - Problem described in Full Detail	Recommended usability solution described in full detail here.	*****

Major findings
(Bulleted or numbered list of major findings)

Major finding #1
[Statement of the major finding]
[Brief summary of finding, may be bulleted list of major data points, often includes a full or partial screen shot with call-out]

Recommendations:
□ May include recommendations for resolving major finding. Recommendations may be a numbered or a bulleted list.

[Each major finding section may also include cross-links to the later pages on specific scenarios that exemplify the major finding]

[Findings should include positives as well as negatives. It is important to let developers know what is working well so that those aspects stay in the product. Also, developers may be more willing to hear about problems when you also let them know about successes.]

Details of the scenarios

Scenario 1.
[Title 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 so well as what is not working. A recommendation may be "Keep this as it is." It may be "Go 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

19

Presentation formats for findings

Topic	Findings	Recommendations	Priority
Flight booking	Most people used Google to find sites and would look at 2 or 3 sites to find the cheapest price before making a booking. One person said they wanted to be sure a site was reputable before making a booking. "These sites assume you are looking not just looking, so you don't get the price as quickly as you would like (UE)"	In view of the perception that Travel site is a --- site, it is important to give users the expectation that fares are ---.	High
Holiday	Secondary navigation: this was not always noticed immediately, but all items on the national variants were understood by the users.	Make secondary navigation more prominent.	Medium
	--- (UK) was understood as an active or adventure holiday and the range of holidays under this heading surprised some people.	Signpost the range of content under ---, particularly singles holidays.	Medium
	---: many people were uncertain about what --- flights were. One thought it meant the airlines like --- and ---.	Provide an explanation of what --- flights are.	Medium
	Some people expected a search box on the holidays homepage.	A search box on the holiday homepage would simplify navigation.	Medium
Several people wanted to be able to select a range of prices and stars for hotels.	Use tick boxes rather than radio buttons.	Medium	
One person misunderstood the meaning of "++".	Clarify labelling.	Low	
A user with a large family expected to find --- on the homepage, and would like to have been able to search for appropriate ---.	Signpost --- if they are available.	Medium	

The "Reset button" is "OK" in the "RESET" screen

One user thought it is not really clear what it is "reset"

It was not at all obvious how to change location once it had been set. Eventually we spotted "RESET" in black text by scrolling horizontally.

Recommendation 1: Change the term "RESET" to "Change location"

Recommendation 1: Make sure that the "change location" function is easily visible on all screens

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.]
[Next finding for this scenario] [Details for this finding]	[Recommendation]
□ [Perhaps bulleted list to give frequency data or individual quotes]	[There may be more than one recommendation for each finding.] [There may be cross-references to other scenarios.]

20

Presenting Conclusions



3. Usability evaluation



One report included pass-fail criteria
 One listed numbers of problems found,
 for each type

