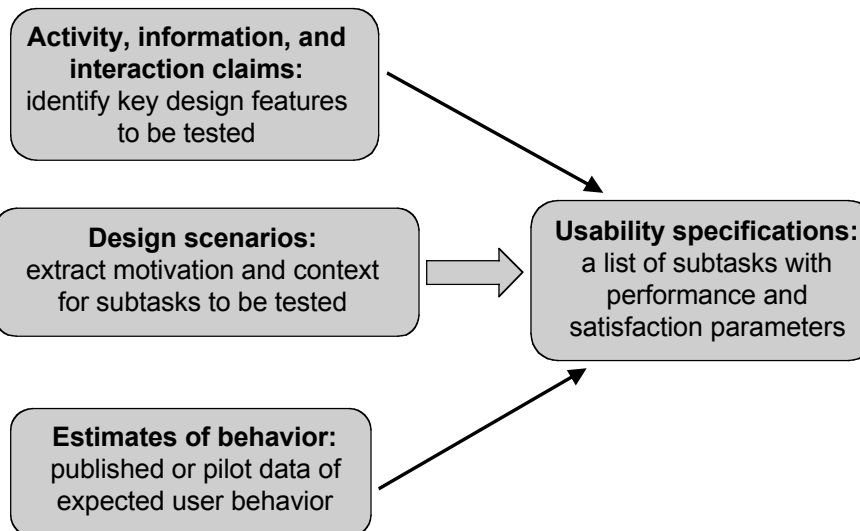


Usability Specifications

- Quality objectives for final system usability
 - like any specification, must be precise
 - managed in parallel with other design specifications
- In SBD, these come from scenarios & claims
 - scenarios are analyzed as series of critical subtasks
 - reflect issues raised and tracked through claims analysis
 - each subtask has one or more measurable outcomes
 - tested repeatedly in development to assess how *well* project is doing (summative) as well as to direct design effort toward *problem* areas (formative)
- Precise specification, but in a context of use

1



2

What about Generality?

- Salient risk in focusing only on design scenarios
 - may optimize for these usage situations
 - the “successful” quality measures then reflect this
- When possible, add contrasting scenarios
 - overlapping subtasks, but different user situations (user category, background, motivation)
 - assess performance satisfaction across scenarios
- Motivation to construct functional prototypes as early as feasible in development cycle

3

Working from a VSF scenario

- (excerpts from the teacher-coaching scenario...)
- when Mr. King meets Sally in the VSF, he can see she is already there, so he selects her name and uses Control+I to see that she is working on her slides, then Control+F to synchronize with her
- he watches her work, and sees her uploading files from her desktop using a familiar Windows browse-file dialog
- when he sees an Excel document, he experiments to see if it is ‘live’, discovers he can edit but not save
- when he sees that she is planning to have visitors come up with their own results using her simulation, he advises her that this will crowd the display, and goes off to find a way to create a ‘nested’ display element
- ...

4

Associated Claims

- Using “Control-I” to identify activities of co-present user
 - + ties info about people directly to their representation on the display
 - + simplifies the screen display by hiding activity information cues
 - but conflicts with the real world strategy of just looking around
 - but this special key combination must be learned
- Exhibit components shown as miniaturized windows
 - + suggests they may contain interactive content
 - but viewers may interpret them as independent applications
- File-browsing dialogs for uploading workstation documents
 - + builds on familiarity with conventional client-server applications
 - + emphasizes a view of exhibits as an integration of other work
 - but the status of these personal files within the VSF may be unclear

5

From Claims to Subtasks

- Identifying and joining co-present users
 - we know that key combinations are harder to learn; how distracting or difficult are they in this case?
- Recognizing and working with active components
 - will users understand these as ‘active’ objects? will they know how to activate them? will they know what is possible when they have done this?
- Importing desktop files into the VSF
 - is the operation intuitive, smooth? is there any resulting confusion about the status of the uploaded files?

6

A Sample Usability Specification

Scenario & Subtasks	Worst Case	Planned	Best Case
Interaction Scenario: Mr. King coaches Sally	2.5 on usefulness, ease of use, and satisfaction	4 on usefulness, ease of use, and satisfaction	5 on usefulness, ease of use, and satisfaction
1. Identify Sally's view and synchronize	1 minutes, 1 error 3 on confusion	30 seconds, 0 error 2 on confusion	10 seconds, 0 error 1 on confusion
2. Upload desktop file from the PC	3 minutes, 2 errors 3 on familiarity	1 minute, 1 error 4 on familiarity	30 seconds, 0 error 5 on familiarity
3. Open, modify, try to save Excel file	2 minutes, 1 error 3 on confidence	1 minute, 0 errors 4.5 on confidence	30 seconds, 0 error 5 on confidence
4. Create nested exhibit component	5 minutes, 3 errors 3 on complexity	1 minute, 1 error 2 on complexity	30 seconds, 0 error 1 on complexity

- Where do targets come from? They are serious, but not absolute
- Notice that we can also “test” overarching scenario.

7

Conducting a Usability Test

- Recruiting of test participants
- Preparation of materials
 - informed consent, background & reaction questionnaires,
general and task-specific instructions, data collection
- Test procedures
 - before, during, after; including assistance policy
- Summarizing and interpreting the results

8

Recruiting Test Participants

- Who are stakeholders, which ones are *actors*?
 - may mean different users for different tasks
 - or, may mean users role-playing other stakeholders
- How do you get people to participate?!
 - participatory design, but this has its own downsides
 - offer them stipends or other rewards
 - make test seem interesting, emphasize novelty
 - last resort, hire from a temp agency...
- Your project: OK to just ask friends, but choose people who can role-play scenario context

9

Informed consent

- Always an issue when human subjects involved
 - the history: psychological research that deliberately deceives people or engages them in moral dilemmas
 - the fix: procedures must be approved by a committee
- Ensures respect for individuals' concerns and hesitations about participating
 - full disclosure of procedures (except when necessary)
 - clear statement of voluntary nature, participant's rights
 - signature indicating understanding and willingness

See p. 256 for a model to use in developing your own informed consent form

10

User Background Questionnaire

- Characterize the user sample you end up with
 - *relevant* experience, expectations, starting attitudes
 - the question: are these the users you need to test?
- But also, helps to interpret test results
 - e.g., experienced computer users will likely do better
 - domain experts may be more critical, more specific
- A range of questions but not too long
 - personal, demographic, experience, current attitudes
 - shoot for one page, seems less intimidating, tedious

See p. 258 for a model to use in developing your own user background survey

11

Developing User Rating Scales

- Convenient for gathering subjective reactions
 - often summarized numerically by mapping judgment categories to ordinal variable (e.g. 1—>5)
 - flexible, can be very general or specific
 - can use to examine opinion change (post-pre)
- Likert scale: measures strength of agreement to an assertion about the system or task domain

Shopping for groceries online is enjoyable.

Strongly
Disagree

Disagree

Neutral

Agree

Strongly
Agree

12

Task Instructions

- General instructions that introduce overall test
- Two sorts of instructions, depending on test type
 - open-ended and goal-directed, for scenario exploration
 - these participants will be doing *think-aloud* process
 - usage context followed by very precise goals, for subtasks
 - clear specification of the user's *goal*
 - avoid options or ambiguities unless part of the test
 - no step-by-step scripts: you are testing the system, not your ability to write complete instructions!

See p. 254 & p. 259 for models to use in developing your own instructions

Task Instructions

Background to Tasks 1-4:

Imagine that you are a neighbor of a high school student (Jeff) who is participating in the VSF. You have discovered that Jeff will be exhibiting his project tonight, and you log on to visit with him. With you is your daughter Erin, who is a middle school student just getting interested in science.

Task 1:

- Find Jeff's exhibit and go to it in the VSF

Task 2:

- Locate Jeff's position in his exhibit, and join him so that you are looking at the same material.

Task 3:

- Join the ongoing conversation (between Jeff and Sarah, another visitor). Let Jeff know that Erin is with you, and ask him to show you around his exhibit.

Task 4:

- Follow Jeff's directions about how to use the "asterisk tool" to mark the three Excel charts as of interest to you and Erin.

Notice that these instructions make assumptions about system state at each point.

Planning for Data Collection

- Be prepared: know in advance *what* and *how*
 - one evaluator interacts with the user, the other keeps track of what happens, collects times, etc.
 - a structured form or template can be very useful
- Take advantage of tools if available and easy to use
 - video taping, screen capture, event logging, etc.
 - particularly useful when collecting think-aloud data
- Know when and how to intervene if necessary
 - a three stage assistance policy: “try again”, “look here”, and finally “just do this: ...”
 - be ready to prompt (“what just happened?”) for users in the exploratory think-aloud condition

15

User Reaction Questionnaire

- Critical for gathering subjective reactions
 - for small tests, interviews can also be even very useful
- Similar in structure to background questionnaire
 - but no demographics this time
 - may include change in opinion due to test experience
 - specific rating scales tied directly to target outcomes in the usability specifications
 - the “three best” or “three worst” features
 - don’t forget the “anything else?” at the end

See p. 261 for model to use in developing your own instructions

16

Task-Specific Usability Judgments

I was confused by commands used to synchronize and un-synchronize with others.
Strongly Disagree Disagree Neutral Agree Strongly Agree

The procedure for uploading files into exhibit components is familiar to me.
Strongly Disagree Disagree Neutral Agree Strongly Agree

Learning that I could not make permanent changes to project data increased my confidence.
Strongly Disagree Disagree Neutral Agree Strongly Agree

Creating a new exhibit element that is nested behind another element is complex.
Strongly Disagree Disagree Neutral Agree Strongly Agree

17

Results: Summarizing User Data

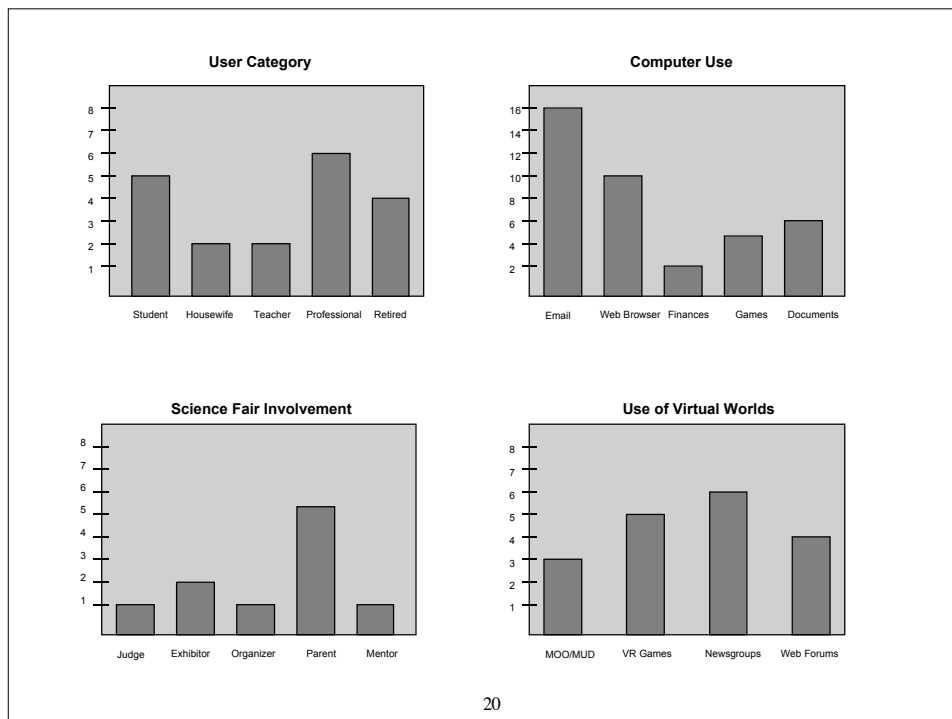
- How you summarize depends on variable type:
 - *categorical*: responses are classified into groups
 - *ordinal*: responses fall in groups, but natural order
 - *interval*: a scale with equidistant values
 - *ratio*: numerical scale with defined zero value
 - *qualitative*: comments to organize and discuss
- Examples of each?
- What are appropriate summary treatments of these differing kinds of variables?

18

Tables for Numerical Data

	Students (N=12)	Community (N=8)	Teachers (N=4)	Administrators (N=2)
Age	14.5 (1.18)	54.7 (7.69)	38.3 (5.41)	45.8 (2.44)
Education	10.0 (1.55)	17.8 (4.43)	16.5 (1.94)	17.3 (3.44)
Residency	10.8 (2.65)	12.3 (3.01)	8.3 (2.87)	5.9 (5.22)

Always be certain to indicate the number of data points in each measure of central tendency (usually referred to as N)



Task Performance Data

VSF Task	Mean Time (seconds)	Mean Errors
1. Synchronizing views	40.2 (3.12)	1.6
2. Uploading a file	115.30 (15.33)	2.2
3. Interacting with Excel	62.55 (2.30)	0.9
4. Adding nested element	143.13 (20.41)	2.8
Combined total	361.18 (10.29)	5.2

*Standard deviation conveys important information about variability
What do these data tell you? What story would you tell? Which tasks would you look at in more detail?*

User Ratings Data

Likert item	Mean rating	Change?
1. VSF is like real world	3.2 (1.55)	+ 1.01 (0.15)
2. Exhibits are diverse, interesting	4.1 (1.67)	+ 0.55 (0.25)
3. Opportunities for science involvement	2.9 (2.11)	+ 0.10 (0.09)
4. Confusion about synchronization	3.5 (0.91)	
5. Familiarity of file uploading dialog	4.2 (1.02)	
6. Confidence when interacting w/ project data	3.7 (2.24)	
7. Complexity of creating a nested element	4.0 (1.94)	

0:00 {User begins marking task.}
"OK, it says that Jeff is going to tell me how to use the marking tool...I guess that means he will be typing something in the chat..."

{User reads aloud Jeff's last message in the chat log}

0:30 {User begins to type something in the message area.}
"The problem is that I don't know what he means by asterisk key! Is there a special button or something that I am missing?"

0:45 {User pauses, presumably waiting for a response from Jeff, then reads}
"Oh, I see what he means, I have to use the keyboard. That's weird, everything else is just mouse and point. Where is the asterisk key anyway?"

1:10 {User is examining keyboard, seems to find asterisk key, returns to read the instructions in the chat, but must scroll up to see them}

1:45 {User points at a Chart on the bulletin board map with mouse, then presses Ctrl+asterisk}
"Very cute. I like the little yellow asterisk. I wonder if I can turn it off by doing it again in place like selection {tries this and it works} ...cool. So now I can go around marking anything I want. I hope it doesn't show to Jeff, though..."

2:20 {User repeats the marking with the other two charts}
"OK, I'm done. What next?"

23

Making Sense of the Results

- Relate to high-level goals: is the system useful, easy to learn and use, satisfying?
 - which of these is hardest to judge in lab study?
- But also compare directly to usability specs:
 - did you miss, meet, or surpass the target level?
 - more importantly, can you figure out why?
- Guidance on how to change design comes from the *details* of the testing, not the summary values
 - why was user confused (or not), why was an interaction difficult or awkward, etc.

24