questionnaires, interviews, observations & planning evaluations

W03 pre-reading quiz results

1. 87.9% got the correct answer
2. 60.6%
   - 30.3% chose all of ‘a’ ‘b’ and ‘c’
3. 88.6%
4. 90.9%
5. 92.4%
6. 72%
   - 3 point Likert and 5 point Semantic Differential were also popular choices
today

- part I: questionnaires
- part II: interviews
- part III: observation
- part IV: planning evaluations

part 1: questionnaires
questionnaires
where else we’re covering it

by now (W03 pre-reading):
• questionnaire structure
• question and response formats
• administering questionnaires

upcoming:
• mini project: you can use a questionnaire in your evaluation

learning goals: questionnaires

• explain when and why questionnaires may be appropriate evaluation technique choice; discuss their pros & cons
• list different styles of questions (open, closed, Likert, etc.) and give examples of what they are appropriate for;
• discuss important considerations for designing and administering a questionnaire
questionnaires: what?

a **tool** in your evaluation toolkit

**definition:** a series of questions used in gathering information from people, usually answered without the presence of a researcher

User Interface Design Process: Evolving Iterations

K MacLean - derived from version by Saul Greenberg (U Calgary)
questionnaires: when and why?

valuable throughout design process

BUT – may be executed differently depending on stage in process:

• **pre/early design:**
  – for understanding
  – good for reaching lots of people early on

• **mid/late design:** input on your design approach and details (prototype, alpha/beta systems, and beyond)
questionnaires: pros & cons

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>administration is cheap - can reach a wide subject group (e.g. mail or email)</td>
<td>creation can be “expensive” - very important to get questions right</td>
</tr>
<tr>
<td>does not require presence of researcher</td>
<td>risk of low response rate</td>
</tr>
<tr>
<td>many results can be quantified</td>
<td>risk of low quality responses</td>
</tr>
</tbody>
</table>

question and response formats

open-ended

closed

• multiple choice, check boxes and ranges
• rating scales
  – Likert
  – semantic differential
• ranked
supplemental notes: question and response formats

For your reference, additional slides covering question and response formats will be included in lecture slides posted online after class.

These were covered in your pre-readings.

question format: ranked

- respondent places an ordering on items in a list
- useful to indicate a user's preferences
- forced choice

Rank the following text editors by preference
(1 = most preferred, 4 = least preferred)

_____ Emacs
_____ Vim
_____ Sublime
_____ Atom
question format: combining open-ended & closed questions

- gets specific response, but allows room for user’s opinion

It is easy to recover from mistakes:

<table>
<thead>
<tr>
<th>disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>agree</th>
<th>please explain:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>the undo facility is great!</td>
</tr>
</tbody>
</table>

designing a questionnaire

establish the **purpose** of the questionnaire:
- what information is sought?
- how would you analyze the results?
- what would you do with your analysis?

determine the **audience** you want to reach

⇒ **pilot** before sending it out:
- test the **wording**
- test the **timing**
- test the **validity**
- test the **analysis**
validity

are your questions getting at what you want?

can increase validity by. . .

• piloting (see how people answer)
• triangulation (target hypotheses with multiple questions)
• use previously validated questionnaires (studied extensively to confirm they gather what they intend to gather)

designing good questions

unlike interviews, can’t ask follow-up questions …so it’s extra important to get questions right

a few general guidelines:

• avoid leading questions
• consider how to order questions
  • previous questions can impact responses
  • think about the logical flow of questions – what should come first? what follows naturally?
• be specific and clear about how users should answer
• keep questions short and easy to follow
• avoid ‘double-’ and ‘triple-barreled’ questions
  • e.g., how often have you used the system and what do you like about it?
• avoid ambiguity and too much room for interpretation
trade-offs

questionnaires are limited by length and complexity

⇒ can’t always ask about everything you want to

try to focus questions on what you really want to learn

• a few focused questions more useful than many general ones.
• if the answer is obvious, you probably don’t need to ask it!

but be careful of focusing too much on what you expect to the exclusion of other explanations

activity 1

questionnaire critique and redesign
administering questionnaires

<table>
<thead>
<tr>
<th>Method</th>
<th>Advantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>in-person administration</td>
<td>• requires time to administer, but highest completion rate</td>
</tr>
<tr>
<td>“take home” (conventional)</td>
<td>• often subjects don’t complete / return the questionnaire</td>
</tr>
<tr>
<td>email</td>
<td>• permits subjects to answer on their own time</td>
</tr>
<tr>
<td></td>
<td>• responses may tend to be more free-form</td>
</tr>
<tr>
<td></td>
<td>• attachments may be a problem</td>
</tr>
<tr>
<td></td>
<td>• response rates depend on trust in source</td>
</tr>
<tr>
<td>web-based forms</td>
<td>• standardize formats and responses</td>
</tr>
<tr>
<td></td>
<td>• Java/JavaScript to ensure correct / complete</td>
</tr>
</tbody>
</table>

be considerate of your respondents and the context you access them in

not just because it’s nice, but it works better.

questionnaire length (short is good):
• think in terms of reasonable completion times rather than number of questions
• do not ask questions whose answers you will not use!

privacy invasions: be careful how / what you ask

ability: limitations like literacy and disability can come into play

motivation:
• why should the respondent bother?
• usually need to offer something in return …but be careful about introducing bias
summary: questionnaires

1. establish purpose
2. determine audience
3. variety of administration methods
   (for different audiences)
4. design questions:
   • many kinds, depend on what you want to learn
   • most important distinction: open/closed (like
     structured/unstructured interview questions)
5. be considerate of your respondents
6. motivate your respondents (without biasing them)

part II: interviews
interviews: 
where else we’re covering it

by now (W03 pre-reading)
• **types:** structured, semi-structured, & unstructured
• **questions:** open/closed, guidelines for good ones
• planning and conducting + recording methods

workshop + upcoming assignments
• practice planning, conducting interviews

learning goals: interviews

• explain **when and why** interviews may be appropriate evaluation technique choice
• discuss pros & cons of interviewing
• outline criteria for a **good** interview, and things you want to avoid doing
interviews: what?

another tool in your evaluation toolkit

involves an interviewer asking one or more interviewees a set of questions, which may be highly structured or unstructured

“conversation with a purpose”
interviews: when and why?

valuable throughout design process

are well suited for:
- exploring issues
- learning more about tasks, scenarios of use
- involving users (+ making them feel involved)
- getting inside the user’s head
- …among other things
interviews: pros & cons

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>excellent for pursuing specific issues</td>
<td>time consuming to conduct and to analyze</td>
</tr>
<tr>
<td>more flexible than questionnaires - probe more deeply on interesting issues as they arise</td>
<td>interviewer can bias the interview</td>
</tr>
<tr>
<td></td>
<td>what interviewees say they do != what they actually do</td>
</tr>
<tr>
<td></td>
<td>may misremember or may not want to tell the truth. If hypothetical question, can be hard to imagine</td>
</tr>
</tbody>
</table>

interviews: infinitely malleable

some things that can vary:

- number of people
  - individual, pairs, groups
- scope
  - duration, depth and breadth
- type
  - structured, semi-structured, unstructured
- location
  - in the lab vs. reality (in context)
- in combination with other techniques
what makes for a good interview?

activity 2
comparing and contrasting interviews

worksheet + video
“How to do a research interview”
some criteria for a good interview

know your goals
• planned questions/interview topics should support your data gathering goals

be organized BEFORE you start
• check equipment
• have all necessary documents (e.g. consent form)

structure the time
• have a clear beginning, middle and end

pilot!
• practice, practice, practice
even very experienced interviewers do this

some criteria for a good interview (cont...)

give participants context
• explain why they’re there, what you hope to learn
  if they don’t know, they can’t tell you

listen
• make eye contact
• refer back to things that have been said
• be attentive, respectful, sympathetic, and flexible
• give the participant time to think
  but if they go off topic, OK to steer them back
some criteria for a good interview (cont...)

use props and visuals
  • combat artificial contexts with props relevant to questions/topics (e.g., prototypes, photos)
  • ask interviewees to bring or draw things to support what they tell you
    sometimes it’s easier to show than to tell
and more . . .
  • this is not an exhaustive list
  • meant to support you in your assignments

let’s practice!

Interview a partner about their TV viewing habits
1. Individually come up with a list of questions to ask your partner [3 minutes]
2. Pair up and determine who is partner ‘A’ and partner ‘B’
3. Partner ‘A’ interview partner ‘B’ [3 minutes]
4. Partner ‘B’ interview partner ‘A’ [3 minutes]
5. Critique each other’s interviews… e.g. any questions that were hard to answer? confusing questions? really interesting questions? [5 minutes]
resources

“how to do a research interview” link to full video:
http://www.youtube.com/watch?v=9t-\_hYjAKww&feature=youtu.be

supplemental notes:
types of interviews

For your reference, additional slides covering types of interviews will be included in lecture slides posted online after class.

These were covered in your pre-readings.
part III: observation

one of the anywhere, anytime evaluation techniques

observation:

where else we’re covering it

by now (W03 pre-readings)
• direct / indirect observation
• in the field, in controlled settings
• strengths/weaknesses

workshop + upcoming assignments
• practice: planning and conducting observation (as part of a larger evaluation); creating a protocol
learning goals: observation

- explain **when and why** observations may be appropriate evaluation technique choice
- discuss pros & cons of different observation techniques
- outline important considerations when planning and conducting an observation and recording data
- describe a contextual inquiry, its relationship to observations/interview, and when it would be an appropriate technique

observation: what?

another **tool** in your evaluation toolkit

**direct observation:** spending time with individuals observing activity as it happens

**indirect observation:** making a record of a user’s activity as it happens to be studied at a later time
Understand USERS:
- who they are
- their key tasks

Understand DESIGN:
- design space and risks
- choose design approach

REFINE Design:
- by element
- considering task
- varied contexts

CONFIRM & debug:
- performance in real use

Early Cycles:
- observation
- interviews
- questionnaires
- contextual inquiry & work modeling

Later Cycles:
- observation
- interviews
- questionnaires
- contextual inquiry & work modeling

User Interface Design Process: Evolving Iterations

K. MacLean - derived from version by Saul Greenberg (U Calgary)
observation: when and why?

valuable throughout design process

BUT – may be executed differently depending on stage in process:
- **pre/early design**: observe for understanding users’ context, tasks, and goals
- **mid/late design**: used to investigate how well the developing design/prototype supports the tasks and goals

observation: pros & cons

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>objective: see what participant does, not says</td>
<td>observer’s presence can disrupt/influence what’s being observed</td>
</tr>
<tr>
<td>flexible: can do in either controlled or real context</td>
<td>can be difficult to analyze or reproduce</td>
</tr>
<tr>
<td>rich data</td>
<td>potentially expensive, time consuming</td>
</tr>
</tbody>
</table>
direct observation (overview)

common approaches:

• simple observation
• think-aloud
• co-discovery learning
• ethnography
• contextual interview (combines interviewing and observation)

simple observation

user is given the task (or not), and evaluator just watches the user

problem: no insight into the user’s decision process or attitude
think aloud method

subjects are asked to say what they are thinking/doing:

• what they believe is happening
• what they are trying to do
• why they took an action

→ gives insight into what the user is thinking

problems

• awkward/uncomfortable for subject (thinking aloud is not normal!)
• “thinking” about it may alter the way people perform their task
• hard to talk when they are concentrating on problem

widely used evaluation method in industry

co-discovery learning

two people work together on a task:

• normal conversation between the two users is monitored
• removes awkwardness of think-aloud, more natural
• provides insights into thinking process of both users
analyzing & interpreting observation data

- qualitative data – interpreted to tell a “story”, categorization and looking for themes
- quantitative data – presented as values, tables, charts and graphs; often treated statistically

depending on what you observe, you might get either kind

… so plan what to record based on what you want to use the data for

More on data analysis next week

Today’s goal: try to obtain data in usable form

coding sheets

goal: try to obtain data in usable form

problem: the observer can introduce subjectivity into observation – where / how can this happen?

what to do? use coding sheets
coding sheets

coding sheets help by …

• adding *structure* and *precision* to observation
• support quantitative analysis \( \rightarrow \) counting events
• tell observer(s) how / what to record
• minimizes subjective contamination

only works if coding sheet is appropriate
important to pilot!

free-form notes can also be helpful

activity 3a

what makes a good coding sheet?
activity 3a: what makes a good coding sheet?

with your neighbour:

1) analyze and discuss the examples [handout]
   • differences?
   • similarities?
   • what does each support?

2) imagine: You want to conduct a think-aloud observation of someone setting up a new GoPro (wearable wide-angle camera) to record a family event.
   • what could you include on the coding sheet?
what influences coding sheet design

1. what data do you need to collect?
   • need goals for your evaluation!
   • goals depend on stage of design!

2. how will you get that data?
   • which observation method you use?

recording observations
(coding sheets work for all of these)

direct/in real time: paper & pencil, typing
   • primitive but cheap
   • evaluators record events, interpretations, and extraneous observations
   • problems:
     – evaluator seems disengaged
     – writing/typing is slow
       – prepared coding schemes can help; just tick off events

audio recording
   • capture discussion (think aloud, co-discovery)
   • hard to synchronize streams (e.g., interface actions with audio)
   • transcription is slow and difficult!
recording observations
(coding sheets work for all of these)

video recording:
• can see what a user is doing
  (good to use one camera for screen + one for subject)
• can be intrusive (at least initially)
• analysis can be challenging
  – takes even longer than audio

companies often build “usability labs” with one-way mirrors, video cameras, etc.
…come visit ours some time (ICICS x7 floor)
activity 3b

conduct and code a think-aloud observation

here’s a task: making origami

Instructions for the task are an interface.

They vary in usability – and different styles work for different people, experience, …
activity 3b: scenario

You are a designer of origami instructions and you want to improve the usability of a set of instructions.

You are going to evaluate these origami instructions by observing participant(s) trying to follow them in a think-aloud observation.

You should assume that the participant has not followed these instructions before. The participant's task will be to use the instructions to complete as much of the origami as possible in the allotted time.

activity 3b: steps

1. Make a draft coding sheet (can use handout)
   - feel free to chat with your neighbour
2. Get into groups 4 - 5. In your group:
   - discuss your individual ideas and design a final coding sheet
   - create at least 2 copies of this final design
3. Conduct the think-aloud observation
   - nominate one person in group to be 'participant'
   - min 2 people: make observations on coding sheet
   - everyone else: take notes on paper
regroup

- what was good/bad? hard/easy?
- what worked better – coding sheets or free-form notes?
- what sort of data did you collect?
observation: summary

- observation: 4 main kinds
- coding sheets: what they are, starting to have some idea of how to use them. [upcoming lecture: using their data]
- observing: what that looks/feels like to do. Things that can go wrong, how to make them go right.

part IV: planning evaluations
planning evaluations: where else we’re covering it

by now (W3 pre-readings)
• advantages/disadvantages of techniques
• methods for recording data

upcoming workshop + assignments
• mini-project: planning an observation + interview
• developing protocols

learning goals: planning evaluations

• develop focus and goals for pre-design activities
  – (e.g. identify human activity needing support, stakeholders, central tasks)
• explain what it means to triangulate in data gathering and evaluation
• make and justify strategic decisions in evaluation planning
example (to consider on your own)
of “understanding the problem”

*Kobo is interested in developing a tablet and line of ‘learning-to-read’ apps targeted at children (and parents). Kobo already has a line of tablets and e-readers, but have not previously studied how usable or successful these products are with children.*

• what is the human activity that needs support?
• who are the stakeholders?
• what are the central tasks?
• what might we want to learn? (evaluation goals)
• who should our participants be?

understanding is iterative

*Pre design: part of evaluation should be to reveal new issues/things you hadn’t thought of*

• early findings inform new rounds of data gathering

what you’re trying to understand generally becomes increasingly specific

– e.g., use unstructured interviews reveal problems, then semi-structured interviews to follow up
– mini-project: only one iteration on this

but works within an evaluation too!

– e.g., use observations to reveal problems, then follow-up with retrospective interview questions
how to design an evaluation

*what should I actually do?*

**one approach:**

*work **backwards** from goals to to figure out what you need to do.*

1. **goals**
2. **evaluation questions**
3. **kinds of data**
4. **type of evaluation**

THEN plan study, collect and analyze data, answer questions…
side note:  
many different levels of goals

over the course of a design process, you may have many goals, at many levels

e.g., scenario of Kobo e reader for kids:

– **Highest-level goal (example):**
  Design an e-reader/tablet that’s well suited to kids

– **Possible (pre-design) evaluation goals:**
  Are the existing reading apps engaging for kids?
  Which existing features of existing e-reader do kids struggle to use?

how to evaluate it?

**evaluation goals:**

– Are existing reading apps engaging for kids?
– Which existing features do kids struggle to use?

1st : what sorts of **questions** get at these **goals**?
• how long do kids spend reading with an e-reader compared to regular books (before getting bored)?
• how do kids feel about using an e-reader?
• how often do kids get stuck when using the e-reader?
• how often do kids need to ask parents for help?

• lots of others possibilities…
how to evaluate it?

2\textsuperscript{nd}: what types of data would answer these questions?
- how long do kids spend reading with an e-reader compared to regular books (before getting bored)?
  \textit{time, magnitude}
- how do kids feel about using an e-reader?
  \textit{opinion, preferences}
- how often do kids get stuck when using the e-reader?
  \textit{counts}
- how often do kids need to ask parents for help?
  \textit{comprehension, counts}
- do families that own e-readers purchase e-books to read with their kids? how many?
  \textit{artifacts}

how to evaluate it?

3\textsuperscript{rd}: what evaluations could give you these types of data?
- counts \textit{controlled observation}
- time spent \textit{controlled observation, data logs}
- stories \textit{unstructured interviews, diaries}
- preferences \textit{interviews, questionnaires}
- opinions \textit{interviews, questionnaires}
- though process/comprehension \textit{interviews, think-aloud}
- artifacts/objects \textit{field observation, interviews}

\rightarrow other methods/combinations possible…
how to evaluate it?

with this info, you should be able to prioritize:

• what specific questions you want to focus on
• and what methods you will use to answer them

then, you’re well set to make a specific plan –
e.g.,
- decide what to include in a coding sheet
  - create interview questions to ask
  - define a protocol for running the evaluation from start to end
  - plan your data analysis

choosing and combining eval methods
depends on goals, questions, & constraints

your need for control over:
  – realism (will results apply in real world?)
  – generalizability (how well will results apply to other situations?)

other things we’ve talked about (or you’ve read in text):
  – natural vs. artificial setting
  – disruptive vs. non-disruptive approaches
  – time, cost, expertise, or resources available
  – stage of development when evaluation is performed

future class: types of data that will result from this
triangulation

a strategy to enhance validity, credibility:
use the *multiple perspectives* available from complementary sources

- interview
- observation
- questionnaire

TRUTH(?)

Use multiple:
- data sources
  - people, places, times
- data collection methods
- researchers/evaluators

mini project overview

4 week mini project: **understand the problems**
with an existing system in terms of the human activity it is meant to support, users, and tasks.

- teams assigned tomorrow in workshop
- topics:
  - see topic options posted on Deliverables page of course website
  - option to choose your own

→ your group will choose a topic by the end your workshop this week
mini project deliverables

2 very short reports: intended as checkpoints to help get you feedback from teaching staff regularly. Lightly marked, but you will turn something in.

- due week 04: Evaluation Plan (checkpoint)
- due week 05: Piloting & Evaluation Status (checkpoint)

1 two slide update: due week 06: 1 interesting finding + 1 question for analysis (2 slides)

1 full report: due week 07: Task Examples, Analysis, Requirements (report)