

# CS 4249: Selected topics in Evaluation

---

KAN Min-Yen

Week 4

# Outline

---

- Examine a few selected topics in evaluation, mostly to supplement what you already know
- Ethics and IRB
- Questionnaire Construction
- Usability Metrics

# Ghostbusters (1984)

---



Perhaps modeled after [Stanley Milgram's](#) experiment?

# Stanford Prison Experiment



- 24 male students recruited in 1971
- 12 randomly assigned as guards, 12 as prisoners
- Recruited ex-convict to inform staff about the environment



- Created a *functional simulation* of a prison
- Prisoners, guards and outsiders (parents and prison warden) fell into their own roles



# The Institutional Review Board (IRB)



[Home](#) [About Us](#) [Guidelines](#) [IRB Forms](#) [References](#) [Glossary](#) [Contact Us](#)

Welcome to  
Institutional Review Board's website

**Investigators**

- Forms
- Guidelines
- Online Training

**Visitors**  
Find out more [About NUS-IRB](#)

**HUMAN RIGHTS**

Please participate in our

**Q:** You're going to do evaluation with feedback from human participants. Do you need to fill out an IRB application?

NUS Institutional Review Board (NUS-IRB) was established on 1<sup>st</sup> September 2003. This Board (also referred to as the Research Ethics Committee in some countries) will review, approve and monitor the ethical aspects of all NUS research projects that involve human subjects and human tissues/cells/data.

Its main objective is to protect the rights and welfare of human research subjects in research activities conducted by NUS students and researchers.

[www.nus.edu.sg/irb](http://www.nus.edu.sg/irb)

# Protecting Participants

---

- Protect their well-being
  - Syphilis Study at Tuskegee, sponsored by US Gov't
- Protect their privacy
  - Safeguard personal information
  - Limit the collection of data relevant to the study
  - Informed consent, especially of images and video
  - Allow people to skip or withdraw from study
- Viewed by some as safeguarding employees and researchers, not their participants
  - Think about the many pages of medical “sign offs” you have to do when undergoing surgery

# Review Exemption

---

- IRB Full Review: Largely for medical research, social sciences (e.g., psychology)
- For IS/CS research: IRB Exemption
  - Important:** Exemption still requires application!
    - Educational settings research, educational tests or instructional techniques and methods
    - Survey without identifiers
    - Analysis of publicly available data or dataset stored without identifiers

# IRB's stand on student research

---

- **Do all student research require IRB review?**
  - **Yes**, if the student research involves human subjects, records gathered on human subjects, or human tissue, and results in publications or generalizable knowledge.
  - **No**, if the student research is part of a class assignment, or part of the requirement for a degree, and the results will not be published.
- **Is a pilot study considered research?**
  - A “pilot study” of a proposed intervention or survey instrument is part of **research** development and is considered **research** under the US Federal definition.

## IRB's stand on student research (2)

---

- **My research was a class project but the results, if published, would benefit the public. Can I submit a request for a retrospective review?**
  - NUS-IRB does not encourage retrospective review.
  - In the event that your supervisor decided that the results of a completed class project should be published, he/she can submit the application for IRB review.
- **What is considered human subject research?**
  - Regulations define a human subject as a living individual about whom a researcher obtains:
    - data through an intervention or interaction with the individual or
    - identifiable private information

# What do you have to provide?

---

These will need to be part of your project documentation as well!

- Title and Purpose
- Description of Procedures
- Duration
- Risks
- Benefits
- Alternatives to Participation
- Confidentiality
- Costs / Additional Expenses
- Participant's Rights
- Contact Information
- Signature

# Forms of Evaluation

---

- **Formative**
  - Done as data-gathering to refine design
- **Summative**
  - Done to compare end product, judge success
- **Expert Review**
  - Heuristic Evaluation
  - Cognitive Walkthrough
- **Questionnaire**
- **Acceptance Testing**
- **Experimental (next lecture)**

Only going to focus on this evaluation method today

# Garbage In, Garbage Out

11 If a bottle of water costs you three times more, would you continue buying it?

12 If not, what would you do?

I will drink regular tap water

I will drink filtered tap water

Other, please specify

13 Are you aware of the steps used to process a generic bottle of mineral water?

14 Do you believe that bottled water can be more expensive than oil?

[mindspace.wordpress.com/  
2006/11/02/  
questionnaires-that-confuse-  
and-confound](http://mindspace.wordpress.com/2006/11/02/questionnaires-that-confuse-and-confound)


[www.portugal.com/blog/bad-  
survey-design-please-stop](http://www.portugal.com/blog/bad-survey-design-please-stop)



# Standardized Surveys

---

- QUIS (U Maryland) [lap.umd.edu/quis](http://lap.umd.edu/quis)
- System Usability Scale (DEC)
  - 10 statements
- PSSUQ (IBM)
- CSUQ (IBM)
- SUMI (HFRG)
  - 50 items (emotional and control, learnability)
- WAMMI
  - Web site evaluation



We'll take a closer  
look at this one

# System Usability Scale – Brooke (1986)

- “Quick and Dirty” Usability – maybe not dirty
- Just 10 items
- Even questions worded negatively

[hell.meiert.org/core/pdf/sus.pdf](http://hell.meiert.org/core/pdf/sus.pdf)

## Why this design?

- Precise measures will vary for system, so not suitable for cross comparison:
  - Against competitors, predecessors
- Actually piloted 50 questions with two known usable and unusable systems; picked questions that polarized the participants
- Every other question inverted to keep people from not paying attention

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

## Quick Qs: SUS

---

Q1: Does a score of 90 on SUS mean that you're better than 90% of the systems out there?

A1:

Q2: Does using inverted questions (i.e., the even questions in SUS) help in getting valid feedback than using normal positive questions?

A2:

## Quick wins

Sometimes I edit surveys. The big scary stats stuff isn't quite my field, but this is what I do to the words:

- Paragraphs and line breaks  
These are critical for ease of reading. People skim web text, so don't make it dense. Most of what I do when somebody asks me to edit a survey is breaking up the text.
- Don't say please  
Get right to the point in the fewest words possible. Avoid "Please supply your email address so we can contact you". "Email (optional):" will do.
- No maths  
If you must make me think, don't make me add up. Try and re-frame the question.
- Keep it on one page  
**Or few, rationally grouped pages.**
- Keep it short  
**Every extra question raises the dropout rate.** It's probably also worth keeping each page's content above the fold.
- Make sure questions aren't leading (unless they need to be)  
You can get some real gems from those "Any other feedback" style free-text fields. But you mostly get nothing. It can be worth adding some gentle suggestions.

[www.simple-talk.com/blogs/2010/01/11/survey-design-just-try-to-write-clearly](http://www.simple-talk.com/blogs/2010/01/11/survey-design-just-try-to-write-clearly)

# Web Based Surveys

---

- + Data already entered
- + Logic of survey can be automated
- + Error checking
- + Response timing and tracing
- + Randomization
- Investigator not around to answer questions
- Participation levels may be lowered\*
- Possibility of multiple submissions

Conclusion? Use web based surveys unless infeasible

# Surveying: Context Effects

---

- The better you know your respondents, the better your response rate will be.
  - Are they stakeholders?
  - Trading off sample size can be worthwhile
- Contact them in advance
  - To create awareness, assemble opt-in
- Don't send unsolicited emails. Only use opt-in.
- Time your surveys to arrive at the right time
  - Professionals: Avoid weekends and beginning of the week
  - Students: Avoid exams
- Schedule reminders
- Offer incentives

[help.surveymonkey.com/euf/assets/docs/pdf/SmartSurvey.pdf](http://help.surveymonkey.com/euf/assets/docs/pdf/SmartSurvey.pdf)

# Usability Metric Considerations

---

- Depending on the usability objective: the overall purpose and the context
- Just because something can be measured doesn't mean it's relevant
- If it is relevant, then what is the acceptable figure to determine success?
- What is the usefulness of the data to be obtained against its cost in resources?

# Usability Metrics

Adapted from Benyon (2010)

Usability Objective	Effectiveness Measure	Efficiency Measure	Satisfaction Measure
Overall Usability	% tasks completed, % users successfully completing	Time to complete, Time on non-productive actions	Ratings scale, Frequency of use
Expert Users	% hard tasks completed, % relevant features used	Time taken	Ratings of advanced features
First Use		Time on first attempt, Time in help functions	
Casual, Intermittent Use		Time to re-learn functions, # of persistent errors	Frequency of reuse

# Usability Metrics (Wider Scope)

Usability Objective	Effectiveness Measure	Efficiency Measure	Satisfaction Measure
Overall Usability	% tasks completed, % users successfully completing	Time to complete, Time on non-productive actions	Ratings scale, Frequency of use
Engagement (Games / Social Media)	Would you do it again, % champions	% Like, % Comment, % Share	Fun-o-meter (Ratings Scale), Sentiment Analysis
Presence (Virtual Reality)	Impervious to outside stimuli	Relate to physical site, Time to navigate	Realism rating

# Summary

---

- **Ethics considerations of evaluations**
  - Institutional Review Board
  - Student Research
- **Focus on survey instruments**
  - Context of the survey
  - Case study: SUS
- **Metrics for measurements**