

CSC 428F/2514F

HUMAN-COMPUTER INTERACTION

Lecture 15
1997 November 3

INTERFACE EVALUATION: DESIGNING EXPERIMENTS TO ANSWER QUESTIONS

presented by
Ben Barkow

15.1	“BUSINESS” VS. “SCIENTIFIC” DECISIONS.....	2
15.2	WHY AND WHEN TO UNDERTAKE EXPERIMENTS.....	3
15.3	FORMING THE QUESTION.....	4
15.4	COMMUNICATING RESULTS, PT 1.....	4
15.5	CRAFTING THE “TEST”.....	5
15.6	RESPONDENTS.....	5
15.7	COLLECTING DATA	5
15.8	ANALYZING OUTCOMES.....	6
15.9	COMMUNICATING RESULTS, PT 2.....	6

Ronald Baecker
Professor of Computer Science
University of Toronto

1. “SCIENTIFIC” VS. “BUSINESS” DECISIONS

1. In the R&D commercial world, information is collected because it helps make somebody’s decision easier.

You need to establish the most likely estimate of the practical state of affairs and address leading issues only. The *cost* of accessing the information is itself a consideration relative to the costs of alternative ways of accessing inferior- or superior-grade information or the cost of the risks of proceeding in darkness [Javits Center wayfinding study for NYNEX]

In business, “Descriptive” statistics are used to characterize and summarize your findings and crafted for rhetorical impact — as Mark Twain observed, “there are *lies*, then *darned lies*, and then *statistics*.” Sometimes you want to “describe” how secure your findings are from the threat of having arisen by chance.

2. Scientists try to make a point, not support a decision.

In “scientific” work, you need to pass through a narrow “gate” guarded by many vicious partisans; they will concede you admittance through the gate if you hit them hard in one theory soft-spot, however unimportant and unrepresentative that spot may be.

Part of the “scientific” outlook is to employ statistics which confirm or disconfirm that your findings are highly unlikely to be a result of random fluctuations.; this is classical “Testing” statistics.

3. Communication effectiveness is inherent in business research, and is definitely inside the loop. What sense would it make to uncover some information which does not gain widespread attention, understanding, and acceptance?
4. Some of us hope that “scientific” research will smarten-up and focus on effectively increasing true and beneficial information; it is probably hopeless. Grad students, take careful note.

2. WHY AND WHEN TO UNDERTAKE EXPERIMENTS

For purposes of this course, “experiments” are data collection activities with participants being tortured in contrasting ways to reveal truths (well, perhaps just forcibly detained and queried).

Experiments are devised to answer question(s). When an experiment is the best approach to collecting the information or when an important question can not be answered any other way, then an experiment makes sense.

Not to be forgotten as other reasons for devising experiments,

1. old experimental data may not be quite right and the truth does get moldy over time,
2. it may be easier to test than to find the facts in handbooks and other publications,
3. running tests may allow for serendipitous findings and stimulate creative thoughts (with many ideas coming from the participants), and
4. there may be no more convincing format when it comes time to present findings.

Also, you might find it beneficial to engage in-house participants (including the decision-makers themselves) in experiments because they may find their experiences convincing and engender buy-in... especially when you do one of those Psychology experiments reeking of deception! [Sherif’s “concept formation” paradigm]

3. FORMING THE QUESTION

If you can formulate a good question, your remaining problems are few. Make sure the question is the right one, exactly what is needed to be known to make a decision. Anything else leads you awry.

Do “thought experiments” to test whether all possible answers to your question will illuminate the decision. “Red Team” your question.

Make it a plain and simple question, not influenced by available methods. The methods *follow* the question.

4. COMMUNICATING RESULTS, PT 1

1. Think through what sentence you would like to utter when you are done. Will you need to control how many “but”s you need to add in?
2. Who is the audience? What do they care about? Are they hard to convince? Do they have prior expectations which might influence your approach?
3. Imagine your stats presentation and what graphics will wow your audience... and be sure your design will permit you to fulfill the intention.

5. CRAFTING THE EXPERIMENT

1. The purpose of the experiment is to answer the question.
2. Use techniques you can explain to your mom or dad.
3. Do “thought experiments” to test whether all possible results to your test will illuminate the question. “Red Team” your design.
4. If you are parameter'izing, ensure there is a good fit between the metrics of the experiment and the metrics of the interface. (If you don't know what this means, be sure to come to class.)

6. RESPONDENTS

1. If you need a big sample, you are doing something wrong. Re-think your methods to ensure that fire strikes.
2. Use contrasting groups to establish differences or to establish the absence of differences when these are points you wish to make.
3. The size of the group needed is a function of the diversity within the group.
4. If you wish to utter things about Group X, you need to have a solid (if minimum) size needed to make your point.

7. COLLECTING DATA

“Clean” work is the hallmark of the Master. Strive for simplicity of tests because that will result in robustness of conclusions [the behavioural seismology research].

8. ANALYZING OUTCOMES

1. Don't use techniques you couldn't explain to your mom or dad.
2. Array your findings so the conclusions stand out.
3. All stats are "Descriptive" but some are more descriptive than others. Classical "Testing" stats are an abomination; they rarely provide meaningful "description" except to experts.
4. Use statistical manipulation to shed light on relationships — such as cross-tabs, correlation, and, quite often, Factor Analysis. You are entitled — even expected — to muck-about in the numbers until they yield up their message.

Aim to learn p-values (probabilities) not to simply reject brittle straw-person hypotheses.

5. Correlational findings are generally trustworthy and are meaningful for all but the worst nit-pickers.

9. COMMUNICATING RESULTS, PT 2

1. Study the prose of journal articles carefully. Do the opposite. Avoid "detective story" format. Tell the reader where you are going and why. Use short sentences.
2. Layer your writing with some specialist stuff sent off to appendices. There are few good reasons ever to use footnotes.