Evaluating Testing
The Qualitative Way

Michael Bolton
DevelopSense
http://www.developsense.com
@michaelbolton
michael@developsense.com

Updates
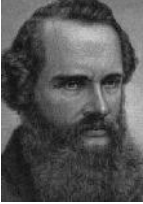
- This presentation is ALWAYS under construction
- Updated slides at http://www.developsense.com/past.html
- All material comes with lifetime free technical support

Introduction

- This is a personal experience report.
- I am not an expert in formal approaches to qualitative research.
- I'm only scratching the surface of the topic here.
- I'm deeply indebted to
  - Cem Kaner
  - James Bach
  - Jerry Weinberg
  - Simon Schaffer
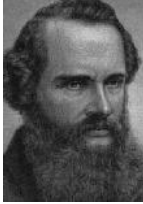 - …and the people on the reading list.

Lord Kelvin

"I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of *science*, whatever the matter may be."
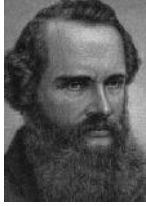
William Thompson, Lord Kelvin

Lord Kelvin

"Heavier than air flight is impossible."
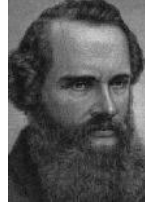
William Thompson, Lord Kelvin

Lord Kelvin

"There is no future for radio."

William Thompson, Lord Kelvin

Lord Kelvin

"In *physical* science the first essential step in the direction of learning any subject is to find principles of numerical reckoning and practicable methods for measuring some quality connected with it." (my emphasis)
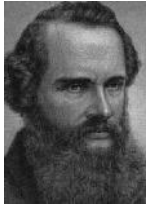
William Thompson, Lord Kelvin

Lord Kelvin

"In *physical* science the first essential step in the direction of learning any subject is to find principles of numerical reckoning and practicable methods for measuring some quality connected with it." (my emphasis)

If you're dealing with things that are important to *people*, and you express things only in terms of numbers, your knowledge is *definitely* of a meager and unsatisfactory kind.

The Physical Sciences

- Physics seeks to know *what will happen* and *what always happens*.
- What will the impact of X be on physical systems?
- Physics is the study of things for which the approximations of physics provide useful results.
- For physics, humans are ideally irrelevant and mostly get in the way of the experiment.

Cem Kaner

- "Software Testing as a Social Science"
- Two talks by Cem Kaner, Waterloo and Toronto, 2006

The Social Sciences

- Social sciences study **humans, in society**
- What will the impact of X be on **people**?
- Use quantitative **and qualitative** research methods
- **High tolerance for ambiguity**, context-specific results
- **Ethics- and values-related** issues are relevant
- **Diversity** of values and interpretations **is normal**
- **Observer bias is an accepted fact of life** and is managed explicitly in well-designed research

"Partial answers that might be useful."

From Cem Kaner, "Software Testing as a Social Science"
http://www.kaner.com/pdfs/KanerSocialScienceSTEP.pdf

Qualitative Research

- A qualitative observation identifies the presence or absence of something
  - "an interpretive issue involving judgment and choice"
  - "meanings rather than frequencies assume paramount significance"
- Qualitative approaches are based on
  - observation
  - making distinctions; categorization and classification
  - description and narration
- Quantitative approaches
  - *assume* that categorizations are accurate
  - largely *ignore* associations with the object of observation

Some Aspects of Qualitative Research

- Putting human values first
- Participant observation
- Storytelling and narration
- Focus on loops of ongoing design
- Skeptical treatment of literature review
- Practices: Concept mapping, memoing, and coding
- Concern for reliability and validity in observation and measurement

A Computer Program

A set of instructions
for a computer.

See the Association for Software Testing's
Black Box Software Testing Foundations course, Cem Kaner & James Bach

A House

A set of building materials,
arranged in the
"House" design pattern.

A House

Something for people to live in.

Kaner's Definition of a Computer Program

- A computer program is
- a *communication*
- among several people
- and computers
- separated over distance and time
- that contains instructions that can be run on a computer.

The purpose of a computer program is
to provide **value** to **people**

Implications of Kaner's Definition

- A computer program is **far more** than its code
- A software product is **far more** than the instructions for the device
- Quality is **far more** than the absence of errors in the code.
- Testing is **far more** than writing code to assert that other code returns some "correct" result

> Quality is value to some person(s).

> Software testing is the investigation of *systems* composed of people, computer programs, products, and the relationships between them.

What Is Testing?

- Excellent testing is not merely a branch of computer science
  - testing *includes* computer science, mathematics, technical domains
  - BUT… focus only on programs and functions, and you leave out questions of *value* and other relationships that include people
- To me, excellent testing is more like *anthropology*—interdisciplinary, systems-focused, investigative, storytelling

| Biology | Archaeology | Language | Culture |

Wade Davis

- Author of *The Serpent and The Rainbow*, *One River*, *The Wayfinders*
- "Dreams from Endangered Cultures", TED Talk, 2007

Central Lessons of Anthropology

- "Every language is an old-growth forest of the mind."
- "Other cultures are not failed attempts to be modern."
- "All of the wisdom of all peoples can contribute to our collective well-being."
- "We think that storytelling can change the world."

> **From Wade Davis, "Dreams from Endangered Cultures"**
> http://www.ted.com/talks/lang/en/wade_davis_on_endangered_cultures.html

To test is to compose, edit, narrate, and justify THREE stories.

A story about the status of the PRODUCT…
  …about how it failed, and how it *might* fail...
  …in ways that matter to your various clients.
A story about HOW YOU TESTED it…
  …how you configured, operated and observed it…
  …about what you haven't tested, yet…
  …and won't test, at all…
A story about how GOOD that testing was…
  …what the risks and costs of testing are…
  …what made testing harder or slower…
  …how testable (or not) the product is…
  …what you need and what
    you recommend.

23

A Healthy and Valuable Testing Mission is Focused on *Learning*

Testers help to defend the value of the product
by *learning* on behalf of our clients.

execution   discovery

design

Exploration
(searching for
value and risk)

investigation

reporting   learning
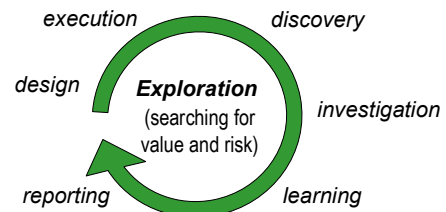
Caution About Literature Review

- Documented knowledge always lags current knowledge in the heads of individuals and communities.
- Not all documentation is of equal relevance or significance; focus is important.
- Your task is not only descriptive, but critical.
- Your conceptual framework is constructed, not found.

- Maxwell, 2005

"Don't restrict your testing to stated requirements! It is a tester's responsibility to validate against *any* legitimate user need we can define, even if the designers are silent about it."
James Bach

Grounded Theory

- Trying to understand something by developing *theories* rooted in *observation*
- A complex and iterative process
- Starts with generative questions that guide research, but that are not static or confining
- Researchers identify core theoretical concepts and examine linkages between them and the data

William V. Trochim, "Research Methods Knowledgebase, 2nd Edition" http://www.socialresearchmethods.net/kb

Testing is grounded theory about a product and its relationship to people!

Some Qualitative Research Practices

- Memoing
  - writing and gathering *memos*—written notes of thoughts and ideas—as the research develops
- Concept mapping
  - mind maps, directed graphs, diagrams, word clouds, sketches, even cartoons
- Coding
  - categorizing observations and describing details and implications

William V. Trochim, "Research Methods Knowledgebase, 2nd Edition" http://www.socialresearchmethods.net/kb

Memos

- Any writing that a researcher does relating to the research, other than field notes, transcriptions or coding
- Personal, reflective, self-critical writing about the research project intended to help you understand it
- Might take the form of journals, notes, exploratory essays
- Be careful of blogs or writing for others; an audience will distort the goals.
- See Maxwell, 2005

A Simple Map of News Reports
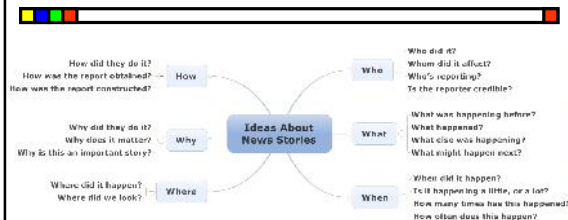
This isn't exactly a checklist, in the sense of "make sure it's there."
It's a *tool* to help us ask questions like "Is this aspect there?"
"What it like?" "How is it different from other reports?"

Questioning News Reports

- Who
  - Who did it?
  - Whom did it affect?
  - Who's reporting?
  - Is the reporter credible?
- What
  - What was happening before?
  - What happened?
  - What else was happening?
  - What might happen next?
- When
  - When did it happen?
  - Is it happening a little, or a lot?
  - How many times has it happened?
  - How often does this happen?
- Where
  - Where did it happen?
  - Where did we look?
- Why
  - Why did they do it?
  - Why does it matter?
  - Why is this an important story?
- How
  - How did they do it?
  - How was the report obtained?
  - How was the report constructed?

Are Test Reports Like News Reports?

- Headline
  - Purpose: to compel attention to the story
- Body
  - the pyramid lead
  - cyclic elaboration of the story
- Narrative
  - how the story follows a coherent path
- Sources
  - what's the relationship between reporter and information?
- Facts
  - what observations are used to back the story?
- Interpretation
  - what the story means for people
  - what is the story's significance
  - are fact and interpretation separated?

More Aspects of News Reports

- Illustrations
  - Are there pictures or diagrams to help with understanding?
  - Are they helpful? Unnecessary?
  - What do they *not* show?
- Background information
  - What knowledge might one need to understand the story?
  - Based on what we know, *what might be missing*?
- Time
  - How does time affect the thoroughness or credibility of the story?
  - How does deadline pressure shape the story?

A Story of Investigating a Bug

- Identifying an effect on people
- Interviewing
- Forming a grounded theory
- Refining research questions
- Engagement with the product
- Manipulating
- Observing
- Literature review
- Getting out of the armchair
- Constructing a narrative

Coding For Exploratory Testing

- Premise: Exploratory testing is NOT just "fooling around with the computer".
- ET is a complex, cognitive activity that includes specific, observable, and trainable activities.
- ET both includes and requires many aspects of tacit knowledge, and activities that generate and refine tacit knowledge.
- Could we create a coding system for exploratory testing?

Exploratory Testing

The Measurement Problem

- Measurement is
  - the application of numbers, based on a model or theory, to attributes of objects or events for the purpose of describing them.
    - Kaner & Bond, 2004
  - the art and science of making reliable (and significant) observations.
    - Weinberg, 1993
- The measurement problem is really the validity problem and the reliability problem.
- It's a more serious problem when we don't recognize that we're using *surrogates*.

Surrogates

- Surrogate *tests*
  - All of our tests model some aspect of the software (typically very simple aspects)
- Surrogate *users* and *business people*
  - Testers are neither users nor business people. When testers can't connect observations to value (and threats to it), they may be ignored.
- Surrogate *measures*
  - Almost all of the quantitative aspects of testing are attempts to *estimate* the quality of something [Kaner, BBST Foundations].

Qualitative Research and Measurement

- The social science research community seems MUCH more concerned with validity and reliability of measurements than the natural sciences are.
- Perhaps this is because good qualitative research seeks more surprises and more controversy.

> "Most of the technology of 'confirmatory' non-qualitative research in both the social and natural sciences is aimed at preventing discovery."
>
> Kirk & Miller, 1985

- Yet even really good quantitative research is aimed towards making observation about *qualities*.

What Makes Stories More Trustworthy? *Validity* and *Reliability*

- Validity: the degree to which you have anticipated and reduced the probability of alternative hypotheses or interpretations
- Reliability: the degree to which your observed results are consistent
  - in different places at the same time
  - in different places over time
  - beware "quixotic reliability"—consistent results because of coincidence and confounding factors
- In qualitative research, reliability has more to do with *credibility, dependability,* and *trust*.

Two Kinds of Validity

- *Construct* validity
  - is this thing we're observing an instance of something in that category?
  - is this (quantitatively) one or zero?
  - is this one of what we're talking about?
- *External* validity:
  - if we observe this over here, can we generalize our description, evaluation, or inference to that over there?

Example: how big is a vehicle?

Now: how big is a "test case"?

Reliability vs. Repetition

- Several generations of scientists have been brought up to believe that repetition (rather than *reliability*) and confirmation (rather than discovery) are the hallmarks of science.
- As testers, our role is to discover *new information*.
- This requires *variation* and *new questions*.

How To Think About Science

Simon Schaffer

- Science has been going through significant changes over the last few decades
- We must look not only at what scientists *say* they do, but what they *really* do
- The "pattern science" is no longer theoretical physics, but the field sciences (e.g. agronomy, field botany)

David Cayley (ed.), Ideas on the Nature of Science
Goose Lane Editions, Fredricton, 2009
Also available as streaming audio on CBC
Search "How To Think About Science"

A Final Thought

If testing's goal is to reveal quality-related information, it would serve us well to sharpen our skills in qualitative research.

Readings

- Mind mapping
  - Tony Buzan
    - *Use Both Sides of Your Brain*
  - Darren McMillan
    - "Mind Mapping 101" http://www.bettertesting.co.uk/content/?p=956
    - "Essential mind mapping: Rapid test design" http://www.bettertesting.co.uk/content/?p=1438
  - Rob Sabourin
    - X Marks the Test Case: Using Mind Maps for Software Design http://www.stickyminds.com/BetterSoftware/magazine.asp?fn=cifea&id=90

Readings

- Qualitative Research
  - Creswell, John W. *Qualitative Inquiry and Research Design: Choosing Among Five Traditions*. Thousand Oaks: Sage Publications, 2007.
  - Kirk, Jerome, and Marc L Miller. *Reliability and Validity in Qualitative Research*. Beverly Hills: Sage Publications, 1986.
  - Maxwell, Joseph Alex. *Qualitative Research Design: an Interactive Approach.* Thousand Oaks, Calif.: Sage Publications, 2005.
  - Stebbins, Robert A. *Exploratory Research in the Social Sciences*. Thousand Oaks, Calif.: Sage Publications, 2001.

Readings

- Anthropology
  - Davis, Wade, K. David Harrison, and Catherine Herbert Howell, eds. *Book of Peoples of the World: A Guide to Cultures*. 2nd ed. National Geographic, 2008.
  - Davis, Wade. *Light at the Edge of the World: A Journey Through the Realm of Vanishing Cultures*. Douglas & McIntyre, 2007.
  - ———. *One River: Explorations and Discoveries in the Amazon Rain Forest*. New York: Touchstone, 1997.
  - ———. *The Wayfinders Why Ancient Wisdom Matters in the Modern World*. Toronto: House of Anansi Press, 2009. http://site.ebrary.com/id/10488286.
  - Fischer, Edward, Ph.D. *Peoples and Cultures of the World*. http://www.thegreatcourses.com/

Readings

- Psychology
  - Ariely, Dan. *Predictably Irrational, Revised and Expanded Edition: The Hidden Forces That Shape Our Decisions*. 1 Exp Rev. Harper Perennial, 2010.
  - Kahneman, Daniel, Paul Slovic, and Amos Tversky, eds. *Judgment Under Uncertainty: Heuristics and Biases*. 1st ed. Cambridge University Press, 1982.
  - Kahneman, Daniel. *Thinking, Fast and Slow*. Penguin, 2011.
  - Levy, David A. *Tools of Critical Thinking: Metathoughts for Psychology*. 2nd ed. Waveland Pr Inc, 2009.

Readings

- Psychology
  - Pinker, Steven. *The Stuff of Thought: Language as a Window into Human Nature*. Penguin (Non-Classics), 2008.
  - Tavris, Carol, and Elliot Aronson. *Mistakes Were Made (But Not by Me): Why We Justify Foolish Beliefs, Bad Decisions, and Hurtful Acts*. Mariner Books, 2008.
  - Weinberg, Gerald M. *The Psychology of Computer Programming: Silver Anniversary Edition*. Dorset House, 1998.

Readings

- Software Testing
  - Beizer, Boris. *Black-Box Testing: Techniques for Functional Testing of Software and Systems*. 1st ed. Wiley, 1995.
  - Copeland, Lee. *A Practitioner's Guide to Software Test Design*. Artech House, 2004.
  - Kaner, Cem, James Bach, and Bret Pettichord. *Lessons Learned in Software Testing*. 1st ed. Wiley, 2001.
  - Kaner, Cem, and Bach, James. 'BBST Foundations' http://www.testingeducation.org/BBST/foundations/.

Readings

- Software Testing and Social Science
  - Kaner, Cem. 'Software Testing as a Social Science' presented at the STEP 2008, Memphis, May 2008. http://www.kaner.com/pdfs/KanerSocialScienceSTEP.pdf.
  - Stevenson, John. 'Are Testers Ethnographic Researchers?' *The Expected Result Was 42. Now What Was the Test?*, January 18, 2011. http://steveo1967.blogspot.com/2011/01/are-testers-ethnographic-researchers.html.
  - ———. 'What You Believe Might Not Be True. (Part 1)'. *The Expected Result Was 42. Now What Was the Test?*, January 26, 2011. http://steveo1967.blogspot.com/2011/01/what-you-believe-might-not-be-true-part.html.