

Deception Dangers of the Numbers Game



LYNN MCKEE

QUEST 2011 Conference

Boston, Massachusetts USA



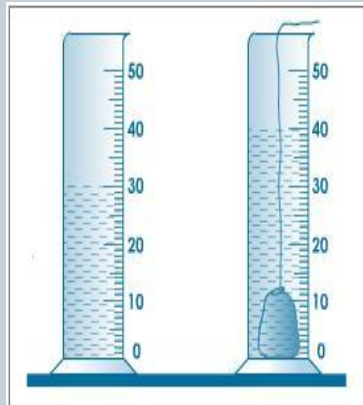
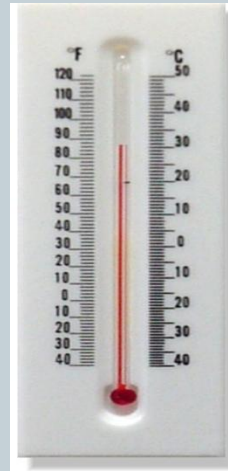
@lynn_mckee #testing #metrics #quest2011

Agenda



- The Measurement Problem
- Measurement & Context
- Qualitative vs. Quantitative Debate
- Positivism & Post Positivism
- Orders of Measurement
- Construct Validity, Abstraction & Aggregation
- Measurement Side Effects
- Inquiry vs. Control
- Summary

What is Measurement?



Measurement



- **What is Measurement?**

- “Measurement is the assignment of numbers to objects or events according to a rule derived from a model or theory.”
- Cem Kaner
- “The art and science of making reliable observation.”
– Gerald M. Weinberg
- Michael Bolton uses Jerry’s definition and adds “Implicit is the notion of comparison for the purpose of making a distinction.”

Why do we Measure?



Track project progress

Gain control of processes

Increase your credibility with your management

Gain the respect of your customers

Gain control of product characteristics

Demonstrate the productivity of your staff

Assess quality levels

Demonstrate the effectiveness of the product

Identify where improvements are needed

Compare different practices

Demonstrate the quality of your work

Determine (relative) complexity or other attributes of the software

- Cem Kaner, Measurement Issues & Software Testing

Common Testing Measurements



Group Exercise



The Measurement Problem



- Distortion and dysfunction are pervasive
- Many things that we would like to measure are subjective
 - complex, qualitative, non-repeatable, and involve human judgment or human performance
- Once you measure something, it changes

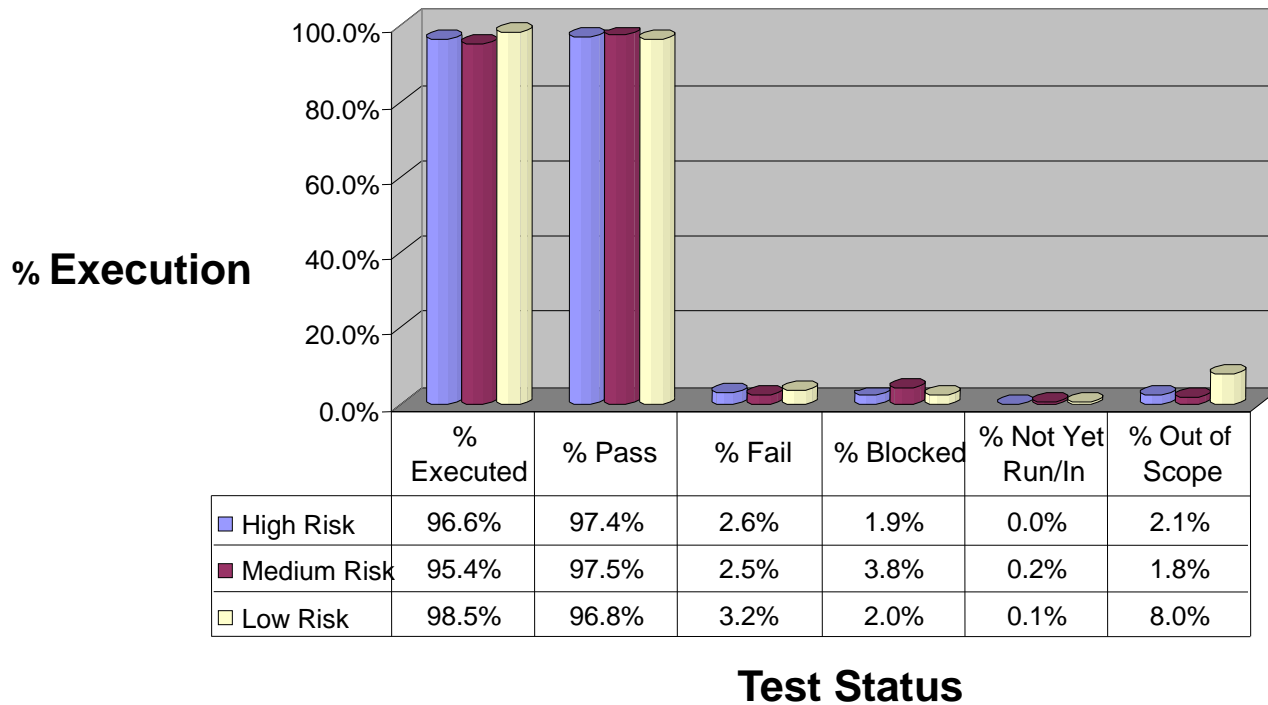
“Decisions about quality are political and emotional, based on discussions and decisions about whose values count and how much they count relative to one another.”

– Gerald M. Weinberg, Quality Software Management, Volume 1: Systems Thinking

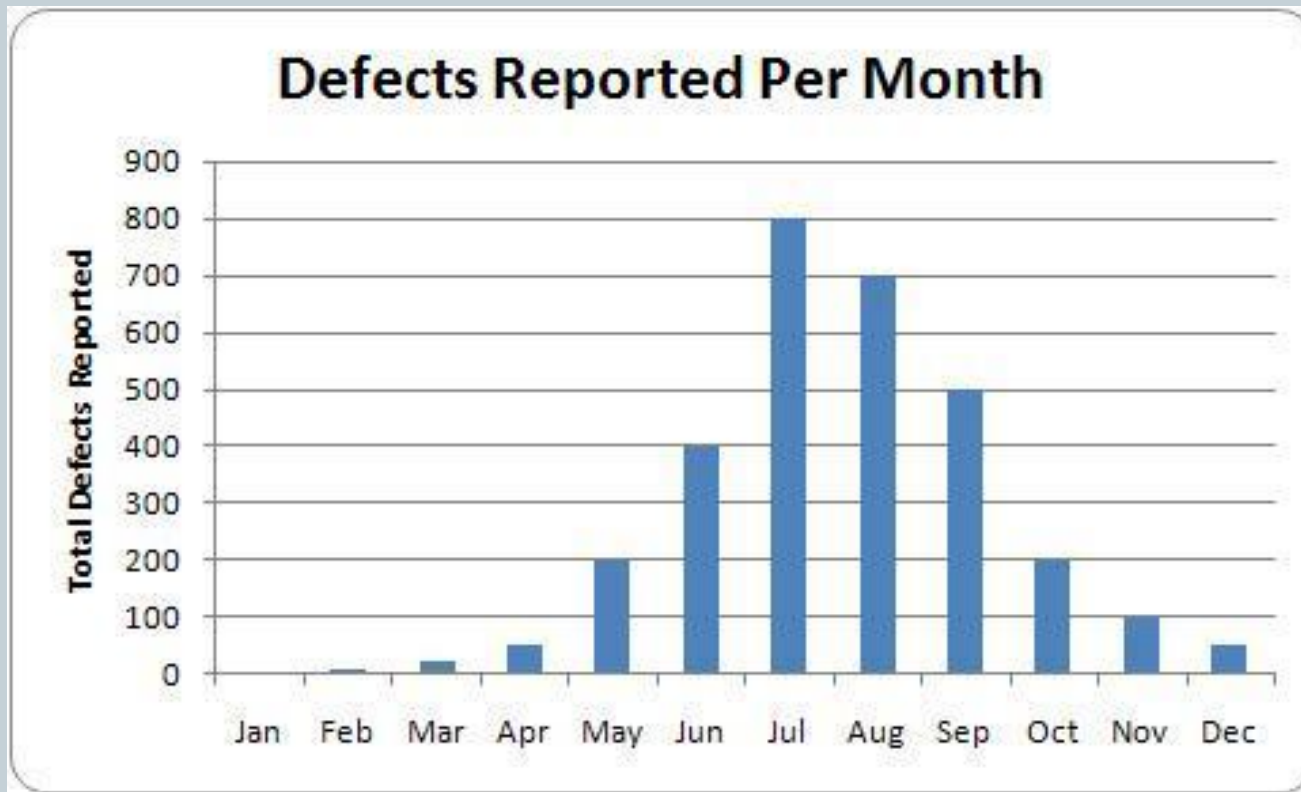
Valuable or Not Valuable?



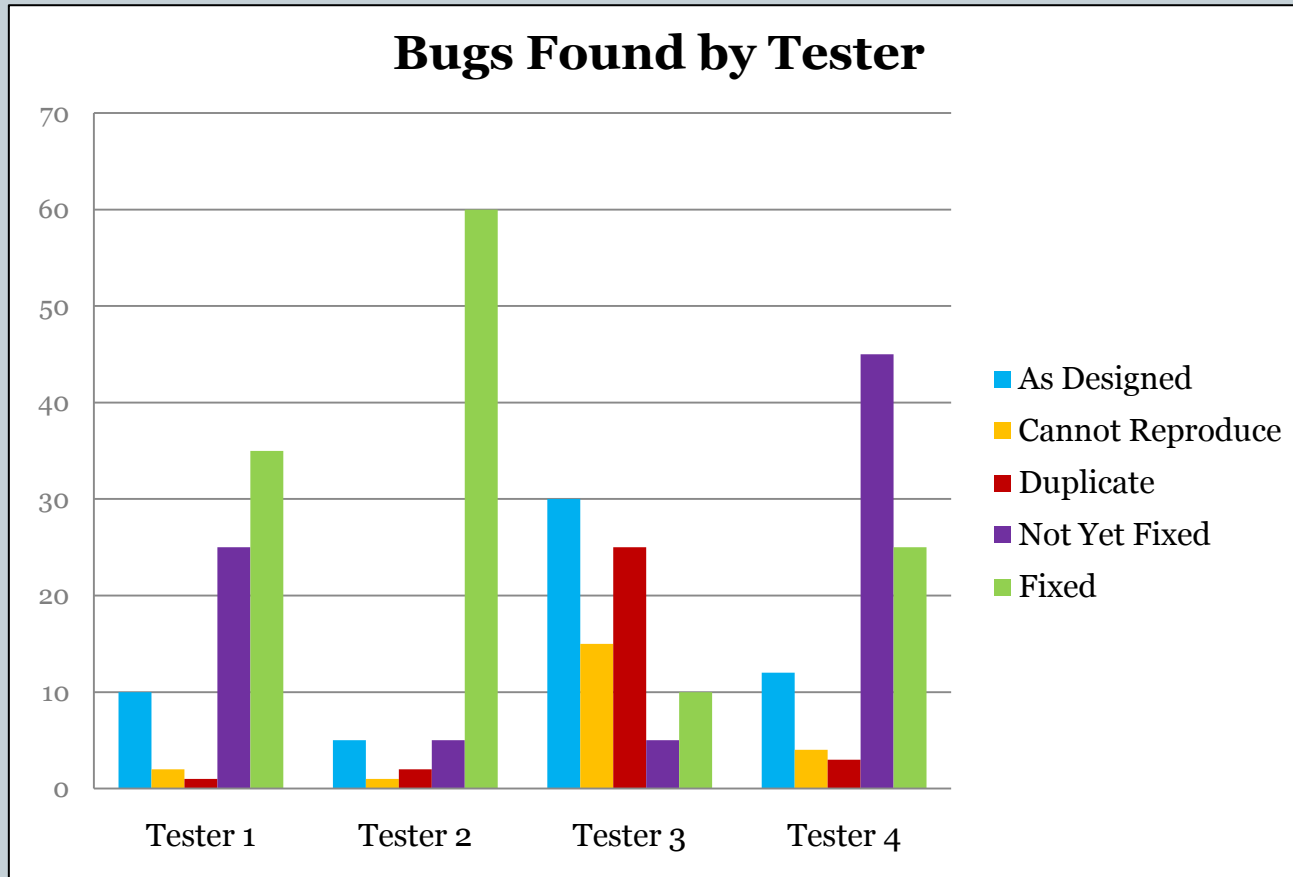
Test Case Execution by Test Status & Risk



Valuable or Not Valuable?



Valuable or Not Valuable?



Measurement & Context



- Measurement without context can lead to ill-informed conclusions
- Be wary of confirmation bias; the tendency for people to favour information that confirms their preconceptions
- Context is important to help clarify ambiguities

“Finding the mean in any given situation is not a mechanical or thoughtless procedure, but requires a full and detailed acquaintance with the circumstances.”

– Aristotle

The Qualitative vs. Quantitative Debate

- **Quantitative = Numbers**
 - Excels at summarizing large amounts of data
 - Reaches generalizations based on statistical projections
- **Qualitative = Words**
 - Excels at "telling the story" from the participant's viewpoint
 - provides the rich descriptive detail that sets quantitative results into their human context

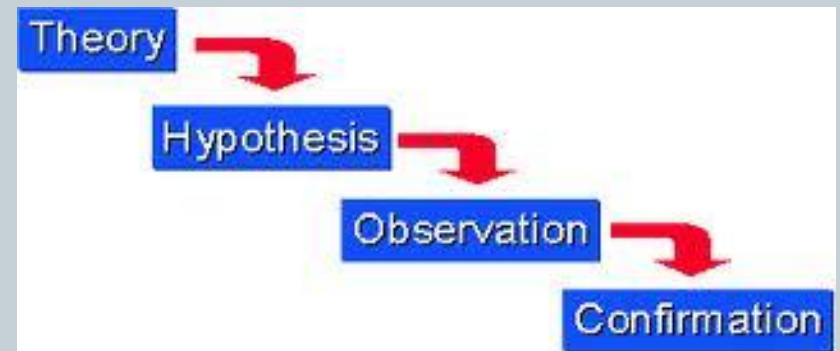


Deductive and Inductive Reasoning

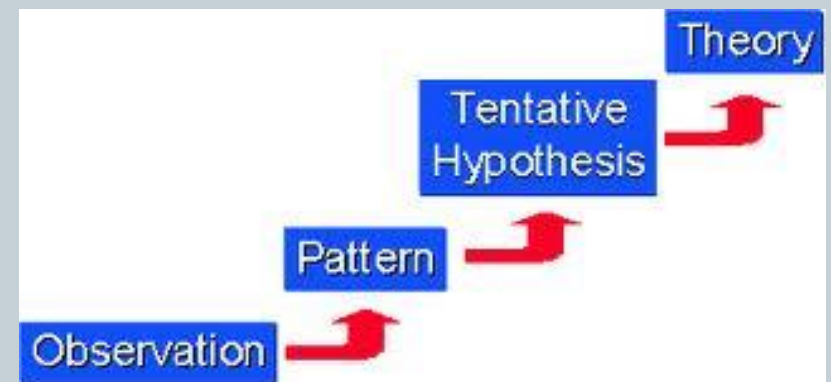


- Common myths:
 - Quantitative is confirmatory and deductive in nature
 - Qualitative is exploratory and inductive in nature
- Consider how your measurements can be gathered through both deductive and inductive reasoning

Deductive



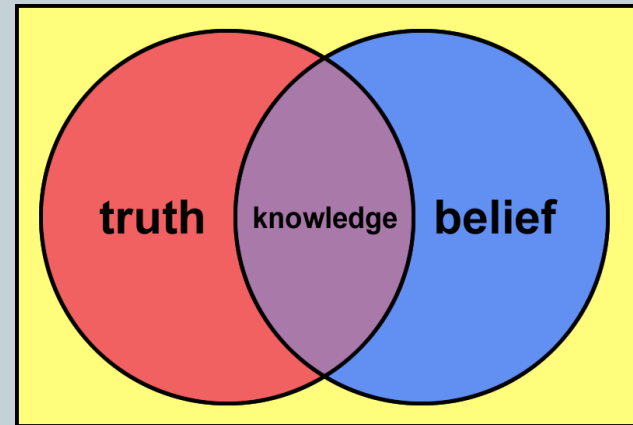
Inductive



Combining Quantitative & Qualitative

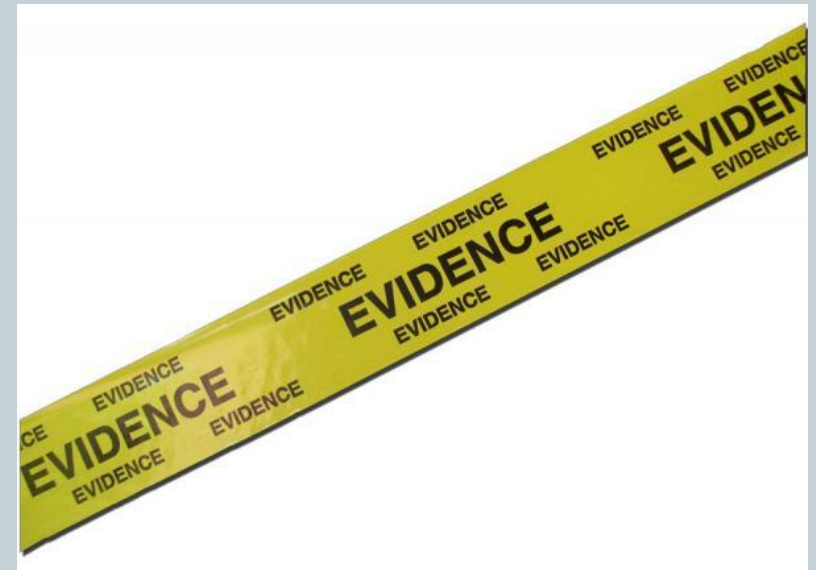


- Numbers alone cannot be interpreted without understanding the underlying assumptions
 - How do we know what we know?
 - What is truth?
 - What are our beliefs?
 - How are we biased?
 - What is reality?
- These assumptions are epistemological and ontological



Positivism Viewpoint

- Rejection of metaphysics, focused only on what we can observe and measure
- Determined to get at truth in order to predict and control
- Believes in *empiricism*; knowledge founded in evidence rather than reasoning, intuition or revelation



Post Positivism Viewpoint

- Rejects the idea that anyone can truly view reality and that all observation is fallible
- Recognizes we construct our view of the world based on our perceptions of it
- Identifies using triangulation across multiple sources to get a handle on reality



Orders of Measurement

- **1st Order Measures**
 - Tend to be qualitative, fast and inexpensive
- **2nd Order Measures**
 - More quantitative, subject to more refined models
- **3rd Order Measures**
 - Precise, high quantitative measures

- Gerald M. Weinberg, Quality Software Management, Vol. 2: First-Order Measurement



Measurement Models & Validity

- How do we know our models are accurate?
- Are the measures meaningful?
- Measurements do not have validity, only our inferences or conclusions can have validity



Construct Validity



Are we *actually* measuring
(are these means a valid
form for measuring)
what (the construct) we
think we are measuring?



Threats to Construct Validity



- There are ten construct validity threats:
 - Inadequate Preoperational Explication of Constructs
 - Mono-Operation Bias
 - Mono-Method Bias
 - Interaction of Different Treatments
 - Interaction of Testing and Treatment
 - Restricted Generalizability Across Constructs
 - Confounding Constructs & Levels of Constructs
 - Hypothesis Guessing
 - Evaluation Apprehension
 - Researcher Expectancies
- For a quick reference guide to construct validity threats see www.socialresearchmethods.net

- Cook and Campbell, D.T. Quasi-Experimentation: Design and Analysis Issues for Field Settings

Abstraction & Aggregation



- Information gain at the expense of information loss
- Abstraction retains only the information relevant for a specific purpose
- Aggregation collects information at the risk of masking specific issues
- We need to question what we are *not* seeing



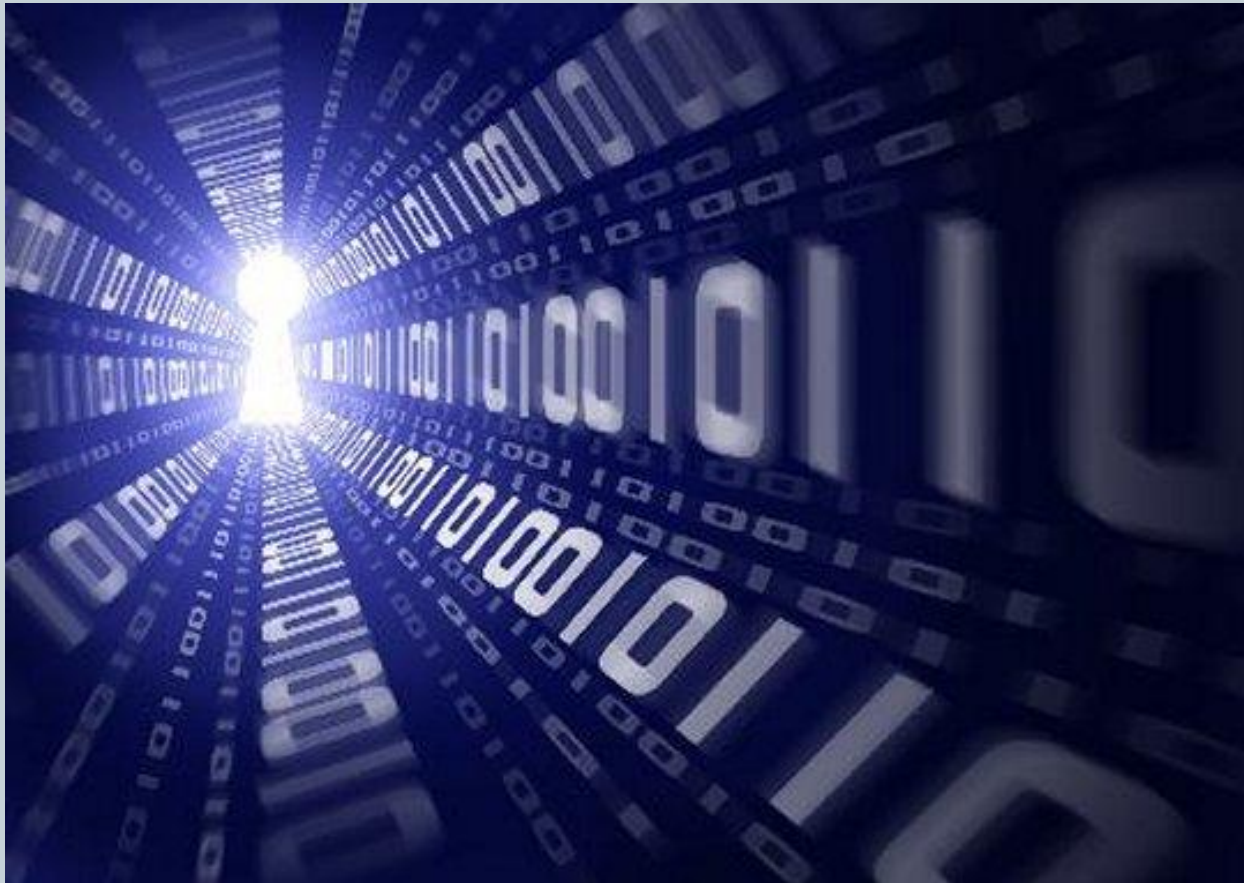
Measurement Side Effects

“Before we can assign numbers to our observations, we must understand the process by which we obtained them in the first place.”

– Gerald M. Weinberg



A Measurement Story...



- Gerald M. Weinberg, Parable of Ones Blogpost

Predictable Behaviour of Measurement

- People tailor their behaviour to things that they are measured against
- Behaviours change in predictable ways to provide the answers management asks for
- The desire to measure the productivity and quality of our work is pervasive... and dangerous



Measures Requested are Measures Attained

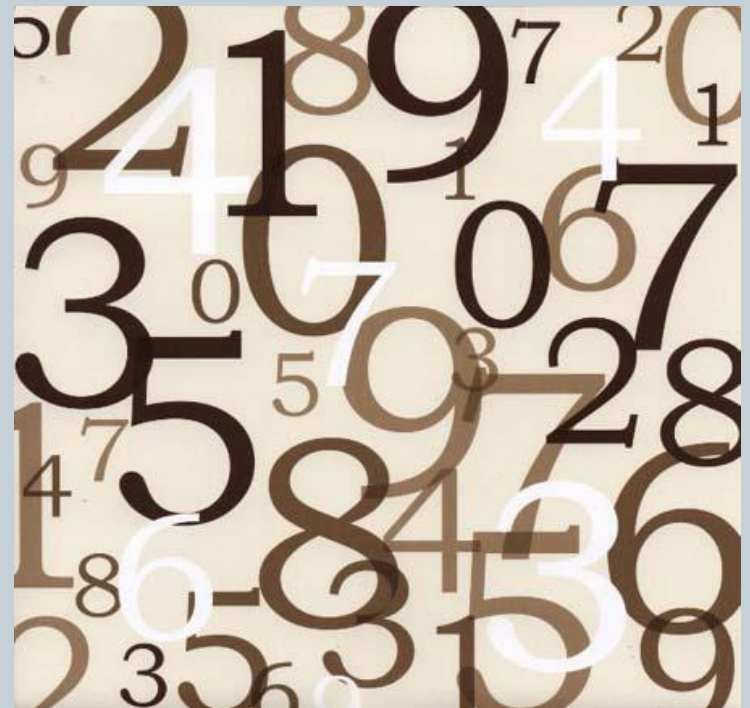


- Consider measuring individuals by their bug reports:
 - If you ask a tester for more bugs, you'll probably get more bugs
 - The measurement system creates incentives for superficial testing and against deep tests for serious underlying errors
 - You can make a tester look good or bad just by choosing what type of testing she should do / what area to test
 - The system also penalizes testers who support other testers
 - Time spent on any process that doesn't lead to more bugs faster is time that counts against the tester
 - People know that other people tailor their behaviour and drives dysfunction within the team

- Doug Hoffman, The Darker Side of Metrics

The Answer is 109 – Good or Bad?

- What question were we really trying to answer?
- We measure to find out if we achieved what we wanted
- Cannot be done effectively with numbers alone as the comparison is qualitative and subjective
- Requires context revealing discussion



Inquiry vs. Control



- Metrics need to be used to drive inquiry rather than to control
- Inquiry explores context
- Use numbers to illustrate stories
- Provide stories with rich descriptive detail that sets quantitative results into their human context



Measurement Inquiry

- Question measurements:
 - Who Says So?
 - How do they know?
 - What's missing?
 - Did somebody change the subject?
 - Does it make sense?

- Darrell Huff,
How to Lie with Statistics



Common Testing Measurements



Group Exercise



Inquiries?



What question were we really trying to answer?

Does this mean we are ready to ship?

What is the story/context?

What coverage model was used?

What was the data sample size?

What other models could we use?

What data is missing?

What are our assumptions?

If bugs are not created equal, does this trend have meaning?

How do we know this to be true?

How has the testing been affected by these measures?

How are these measures affecting the team?

Tell me more...

Does this represent the value of the test effort? How?



Summary



- Measurement is difficult as what we seek to measure is subjective
- Focus on “What question are we really trying to answer?”
- Seek simplicity using 1st and 2nd order measurements
- Question the validity of measurement models
- Use numbers to illustrate stories, watch out for numbers becoming placeholders for stories
- Be critical of our ability to know reality with certainty; consider the epistemological and ontological assumptions
- Think carefully about the potential side effects of measures
- Use metrics to drive inquiry rather than to control

Questions?



Lynn McKee
Quality Perspectives

Calgary, AB Canada

lynnmckee@qualityperspectives.ca

www.qualityperspectives.ca



lmmckee



@lynn_mckee



lynnmckee

References



- Quality Software Management, Vol. 1: Systems Thinking by Gerald M. Weinberg
- Quality Software Management, Vol. 2: First-Order Measurement, by Gerald M. Weinberg
- How to Lie with Statistics by Darrell Huff
- Software Engineering Metrics: What Do They Measure and How Do We Know? by Cem Kaner, <http://www.kaner.com/pdfs/metrics2004.pdf>
- Negligence and Testing Coverage by Cem Kaner, http://www.kaner.com/pdfs/negligence_and_testing_coverage.pdf
- The Impossibility of Complete Testing by Cem Kaner, <http://www.kaner.com/pdfs/imposs.pdf>
- Three Kinds of Measurement and Two Ways to Use them by Michael Bolton, http://www.stickyminds.com/s.asp?F=S15136_ART_2

References



- Measurement Issues & Software Testing by Cem Kaner
http://www.kaner.com/pdfs/measurement_segue.pdf
- The Darker Side of Metrics by Doug Hoffman,
<http://www.softwarequalitymethods.com/Papers/DarkMets%20Paper.pdf>
- Experimental and Quasi-Experimental Designs for Generalized Causal Inference by Shadish, Cook and Campbell
- D.T. Quasi-Experimentation: Design and Analysis Issues for Field Settings by Cook and Campbell
- Parable of Ones by Gerald M. Weinberg
<http://secretsofconsulting.blogspot.com/2010/12/parable-of-ones.html>
- Black Box Software Testing Foundations, Association for Software Testing, <http://www.associationforsoftwaretesting.org>
- Software Metrics: A Rigorous and Practical Approach by Norman E. Fenton and Shari Pfleeger

References



- Aristotle's Ethics, <http://plato.stanford.edu/entries/aristotle-ethics/>
- Social Research Methods, <http://www.socialresearchmethods.net/>
- A Quick Reference Guide to Construct Validity Threats, <http://www.socialresearchmethods.net/tutorial/Driebe/tweb3.htm>