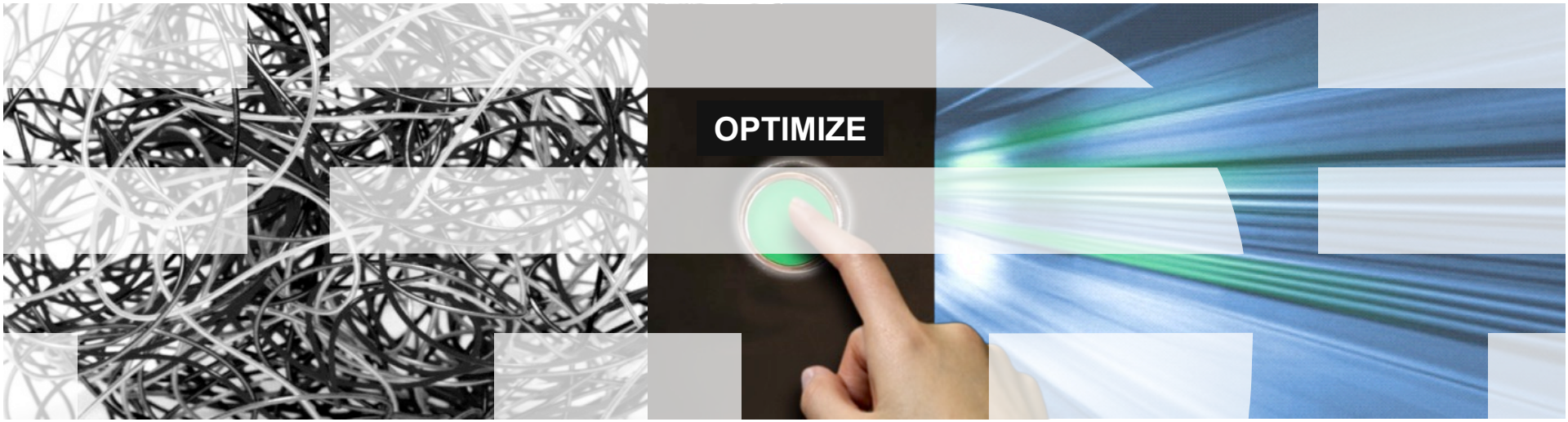
# Behavioral Economics and the Design of Systems Engineering Measures

Barclay Brown,ESEP
Global Solution Executive
IBM Rational
barclay@barclaybrown.com



*In theory there is no difference between theory and practice. In practice there is.*
Yogi Berra

## Difficulties in both theory *and* practice:

- **Why measure?**
- **What should we measure?**
- **How should we measure it?**
- **What does it mean?**
- **What effect will the measurement have?**

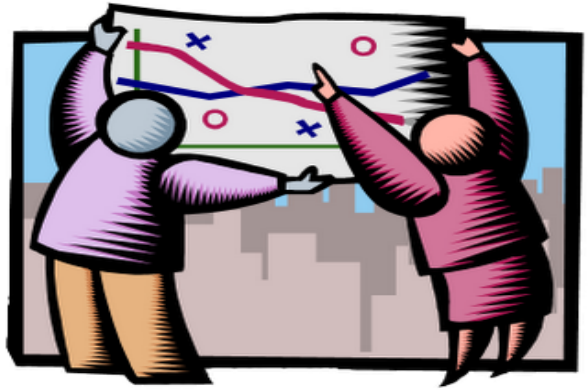
# Why measure?

- Today many projects failures are attributed to poor requirement management and quality
- Metrics are key indicators of project scope, growth, stability, quality and progress.
- Helps engineers and manager's spot trouble before it happens:
  - Enabling early detection of problems that require management attention
  - Providing management visibility into actual versus planned performance
  - Facilitating process improvement with best practice guidance and compliance tracking
  - Gives you continual process improvement with repeatable processes

# What should we measure?

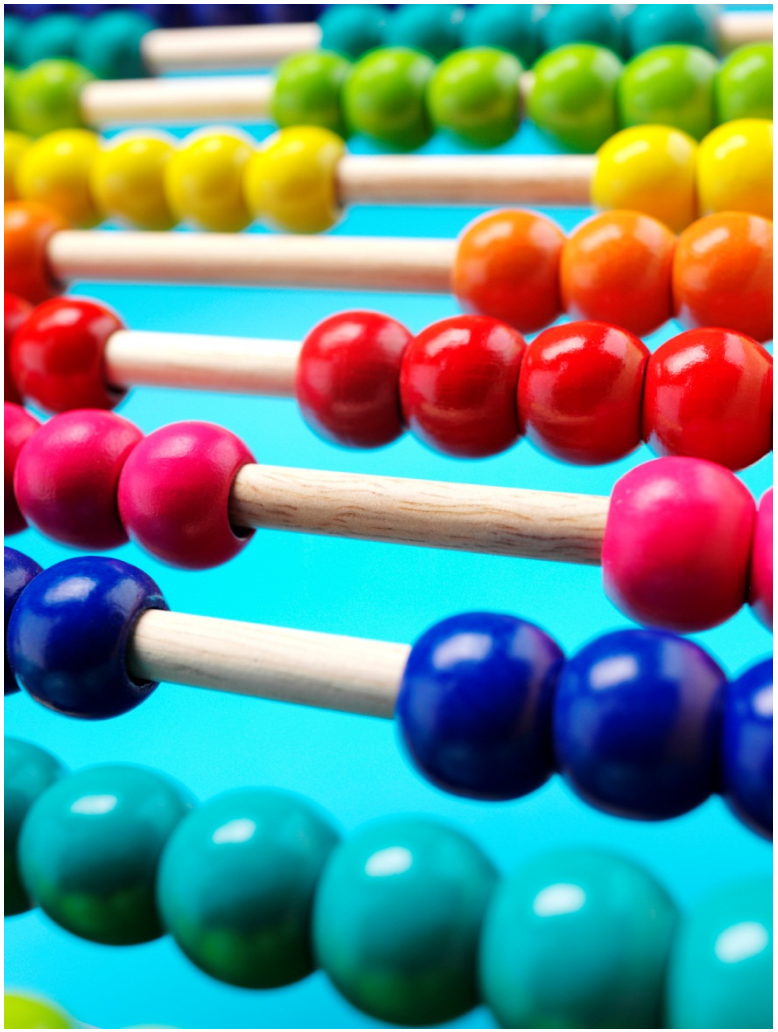
- Aerospace and defense firms 2011 surveyed
  - **Measures that nearly everyone does** in fact measure and use in managing the business. (cost and Schedule)
  - **Measures that some organizations do** (or at least some programs in some organizations) measure, and may use to some extent in managing the business. w/lack of widespread agreement
  - **Measures that are thought to be potentially valuable** but which are not currently being measured because of difficulties in implementing measurement systems. (obstacle is the availability of a measurement system that is reliable, honest, automatic and transparent, lack of standardization)

- What do we need to know? This maybe different then what your managers want to know
  - How close are we to finishing a task or to a major milestone
  - How well is the work being done
  - Do I have the data/information I need to talk to managers so they understand why and what I need to get my job done well?
  - Leading indicators



# Mechanics of Measurement

- Counting things
  - Requirements, defects, changes, model elements, interfaces, work products
- Process Compliance
  - Audit results, work products inspections, review results
- Measures of completion
  - Verification, test cases passed, task completion level
  - “90% done and 50% to go”

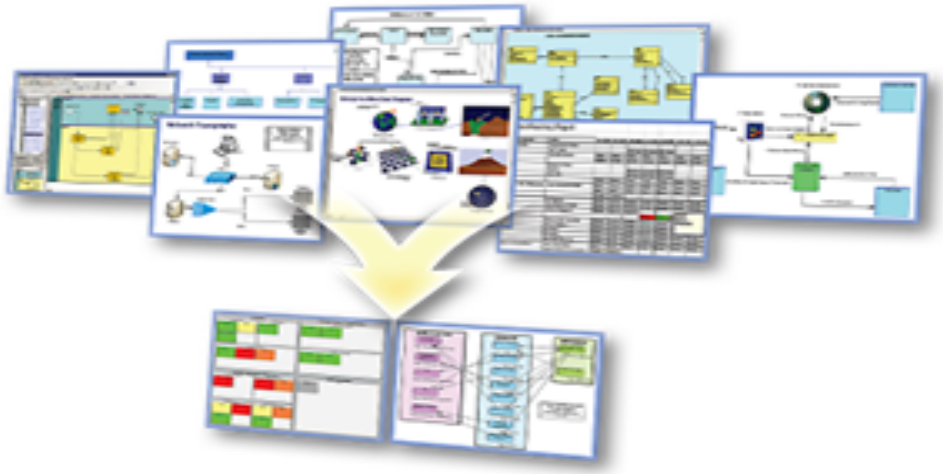


# How should we measure?

- Measuring systems can be challenging
- Measurement systems:
  - Simple manual collection of information from individuals
  - Formal status reporting mechanisms and formats
  - Fully automated systems that collect data in the background without even requiring human input.
- Manual collection
  - Human error
  - **Gaming**
  - Costly
  - Old data
  - Collected using different criteria
- **Automated systems** to collect information can help avoid the expense, time-lag and labor required of manual systems.

# What does it mean?

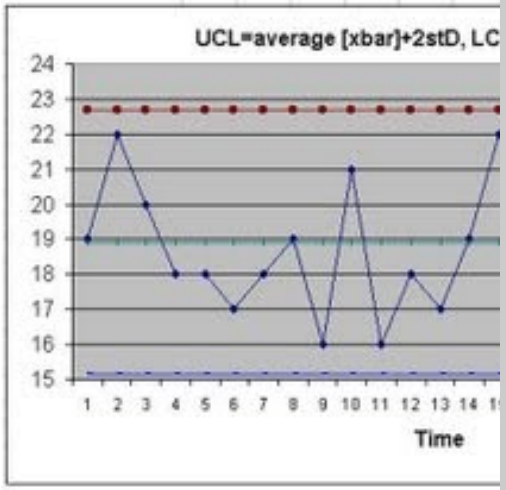
- Deriving Meaning from Data
- Large challenges
  - Even the most obvious measures, such as cost and schedule variance do not always easily translate into business measures.
- In order for measures to be directly meaningful and actionable, three things must accompany each measure:
  - **Norms.** A measure should have a range that is considered normal
  - **Standards.** A measure should have a standard, that is, a minimum-acceptable value
  - **Business Value.** Ultimately measures must track to business impact or business value on the program, expressed as cost or schedule



# Interpreting Measures

- Humans will always add meaning to whatever we see
- We are not good at seeing **randomness** and recognizing it as randomness
- Over-emphasis on interpretation can result in process tampering (Deming)

| Time | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Measurements | 19 | 22 | 20 | 18 | 18 | 17 |
| UCL | 22.70 | 22.70 | 22.70 | 22.70 | 22.70 | 22.70 |
| mean | 18.92 | 18.92 | 18.92 | 18.92 | 18.92 | 18.92 |
| LCL | 15.14 | 15.14 | 15.14 | 15.14 | 15.14 | 15.14 |




**What is the real story?**

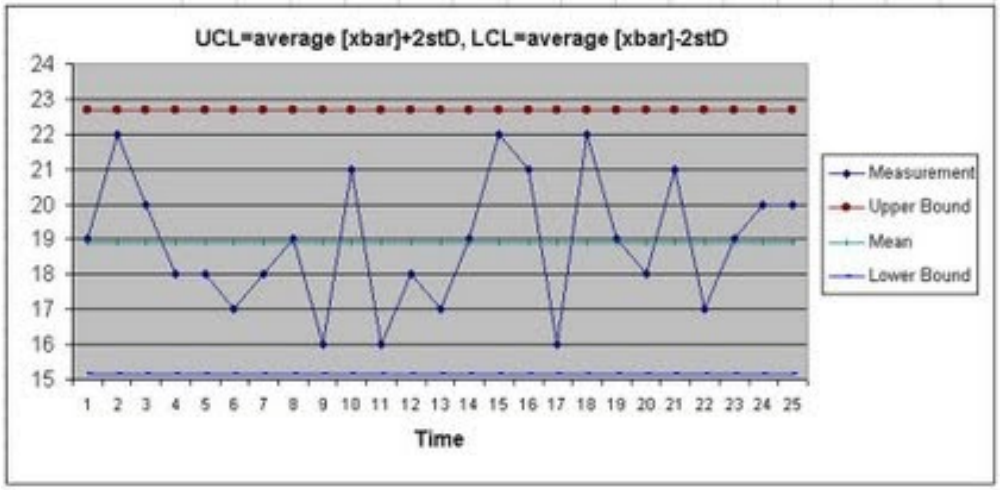
A process analysis run chart, used to determine if a process is in a state of statistical control. Average measurements in process samples versus time. The center line is drawn at the process mean. The upper [UCL] and lower [LCL] control limits are at two standard deviations from the center line.

Source: JF Giraud eng., SPC Examples, https://sites.google.com/site/9264271/spcexamples

# Interpreting Measures

- Humans will always add meaning to whatever we see
- We are not good at seeing randomness and recognizing it as randomness
- Over-emphasis on interpretation can result in process tampering (Deming)

| Time | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Measurements | 19 | 22 | 20 | 18 | 18 | 17 | 18 | 19 | 16 | 21 | 16 | 18 | 17 | 19 | 22 |
| UCL | 22.70 | 22.70 | 22.70 | 22.70 | 22.70 | 22.70 | 22.70 | 22.70 | 22.70 | 22.70 | 22.70 | 22.70 | 22.70 | 22.70 | 22.70 |
| mean | 18.92 | 18.92 | 18.92 | 18.92 | 18.92 | 18.92 | 18.92 | 18.92 | 18.92 | 18.92 | 18.92 | 18.92 | 18.92 | 18.92 | 18.92 |
| LCL | 15.14 | 15.14 | 15.14 | 15.14 | 15.14 | 15.14 | 15.14 | 15.14 | 15.14 | 15.14 | 15.14 | 15.14 | 15.14 | 15.14 | 15.14 |




A process analysis run chart, used to determine if a process is in a state of statistical control. Average measurements in process samples versus time. The center line is drawn at the process mean. The upper [UCL] and lower [LCL] control limits are at two standard deviations from the center line.

Source: JF Giraud eng., SPC Examples, https://sites.google.com/site/9264271/spcexamples

# The Need for Science Experiments

- Hypothesis testing
  - "A requirements volatility that does not decrease by at least 50% after the PDR milestone indicates a lack of sufficient system definition and will cause a high degree of later rework."
- A **study of this hypothesis** would require
  - Historical data of programs' requirements volatility over time
  - Historical data of program's rework
  - Correlation study
- How many such hypotheses are possible? How many studies are needed?

# Behavioral Economics and the Cobra Effect

- Unintended consequences of implementing a change
- Especially, OPPOSITE consequences
- Implementing a measure is like a move in a game
- Consider counter-moves, and rounds of play

| Round | Situational Condition | Game Move | Counter-Move |
|---|---|---|---|
| 1 | Too many cobras loose | Bounty on cobras | Breeding cobras for profit |
| 2 | Too much money spent on bounties for bred snakes | Discontinue bounty | Breeders release cobras |

# More game playing…

- Use look-ahead thinking
- Ask, “if this measure were my ONLY priority, what could I do to make it come out better, even at the expense of other concerns?”
- This is not hard—people naturally think this way!

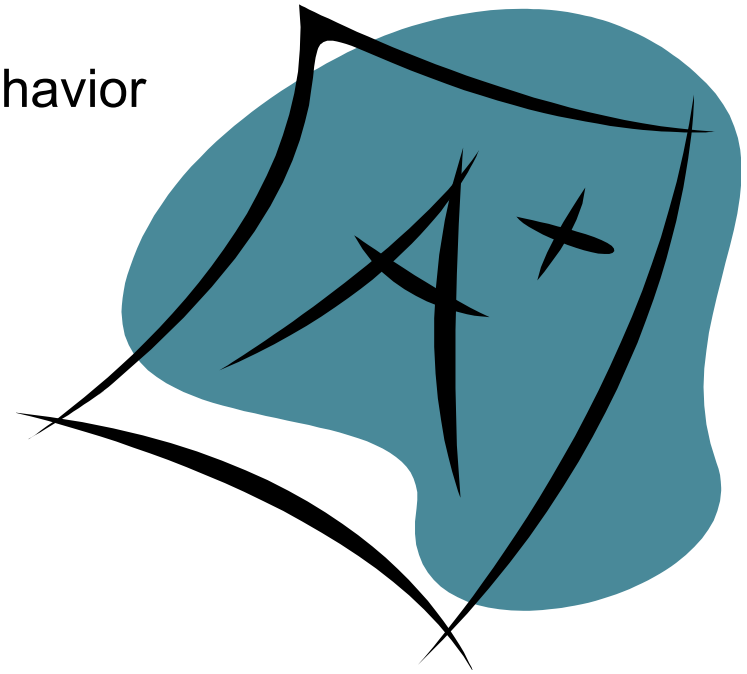
| Round | Situational Condition | Game Move | Counter-Move |
|---|---|---|---|
| 1 | Desire to measure and increase effectiveness of call center **lowering** costs | Measure call completion rates | Calls are marked complete before issue is resolved, resulting in repeated calls |
| 2 | Call volumes increase | Hire more workers, **raising** cost of call center | … |

# Requirements Volatility / Stability

- Little or no ability to use this measure in a meaningful way.
- Ratio of requirements changes to total requirements (Roedler, 2010)
- Baselines and goals must be established
- Not just “lower is better”
- But imagine IF lower was thought to be better,
- What unintended behavioral consequences might result?
- How can this measure be “gamed”?
  - Increase the total number of requirements
  - Decrease the number of new or changed requirements.
- Could cause “batching up” of changes to a requirement and waiting to change a requirement until it must be changed—perhaps just before a major milestone or baseline
- Could cause more requirements to be more out of date than otherwise.

# The Goal: A Perfect Measure

- Measure completely drives the right behavior
- No opportunity for gaming
  - Revenue?
  - Body Mass Index?
  - Grade Point Average?
  - Earned Value?
  - Repeat business?

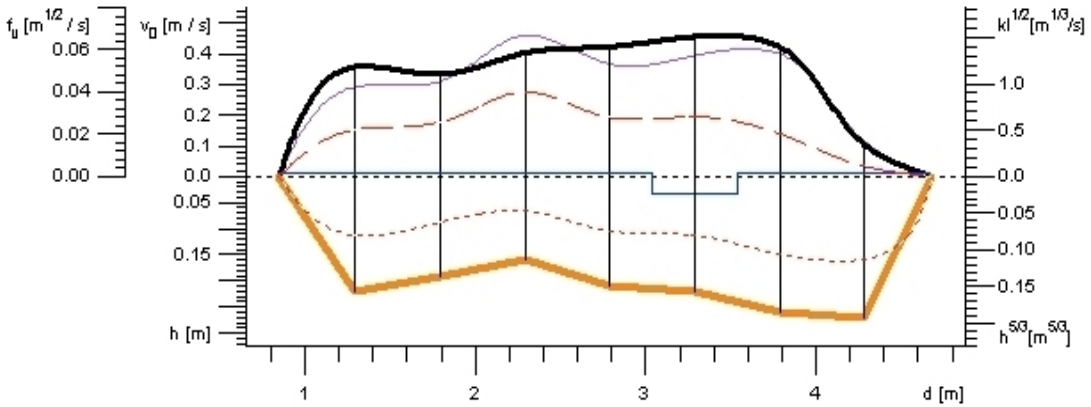


# A Strategy: Balancing Measures

- Implement an additional measure
- Make it just as important
  - Revenue / Profit
  - Listing Price / Selling Price
  - Defect Density / Number of defects corrected

# Leading Indicators

- Leading indicators are a sophisticated form of measurement. You extract some of this data directly from program data and you also need historical data to determine.
- Leading indicators supports managers by:
  - Providing management visibility into actual versus planned performance
  - Gives you continual process improvement with repeatable processes
- The INCOSE Leading Indicators define what they can be and assumes certain tasks are being done on each program in a consistent way. There are 18 defined in the V2 release.




- **3.1 Requirements Trends**
- 3.2 System Definition Change Backlog Trends
- **3.3 Interface Trends**
- **3.4 Requirements Validation Trends**
- **3.5 Requirements Verification Trends**
- 3.6 Work Product Approval Trends
- **3.7 Review Action Closure Trends**
- 3.8 Technology Maturity Trends
- **3.9 Risk Exposure Trends**
- **3.10 Risk Treatment Trends**
- 3.11 Systems Engineering Staffing and Skills Trends
- 3.12 Process Compliance Trends
- 3.13 Technical Measurement Trends
- 3.14 Facility and Equipment Availability
- 3.15 Defect and Error Trends
- 3.16 System Affordability Trends
- 3.17 Architecture Trends
- 3.18 Schedule and Cost Pressure...

# The Path Forward

- Still much work to do in developing comprehensive measurement approaches in engineering organizations.
- Challenges:
  - Choosing measures,
  - Implementing automated systems
- Use scientific hypothesis-testing to establish meaning of measurement results
- Use behavioral economics and game theory to study effects of measures.


**www.ibm.com/software/rational**