Measures and Metrics: The Need for Consistency in HSI Measurement Terminology

LTC STONEY TRENT, PH.D. USCYBERCOM

ROBERT HOFFMAN, PH.D. INSTITUTE FOR HUMAN MACHINE COGNITION
Agenda

• Measurement Pitfalls
• Measurement Theory
• HSI Measurement Example (Cyber Protection Team Technologies)
• Summary
Reliance on traditional “human performance measurement”

→

Failure to measure cognitive work at the systems level

R&D Programs ask for systems that are “adaptive” or “resilient.”

OK. So how do we measure such things?

Step 1: Measure what can be easily measured. OK
Step 2: Disregard that which cannot be measured. Artificial and misleading.
Step 3: Presume the unmeasurable is not important. Blindness.
Step 4: Say the unmeasurable doesn’t exist. Suicide.

Measurement Terminology

**Theoretical Concepts**
- Things or phenomena you would like to understand

**Measures**
- Things you can measure and evaluate

**Operational Definitions**
- Replicable measurement procedures

**Measurements**
- Values associated to events

**Measurement scale**
- Conceptual and mathematical relationships of measures

**Metrics**
- Thresholds or benchmarks for an evaluation

Military Operational Assessments...
- Include these...

...confuse these...
...and rarely consider these
# Measurement Scales

<table>
<thead>
<tr>
<th>Qualitative (Nonparametric)</th>
<th>Quantitative (Parametric)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Nominal – Categories (Colors)</td>
<td>• Interval – Meaningful distances (Time)</td>
</tr>
<tr>
<td>• Ordinal – Ordered Categories (Sequence)</td>
<td>• Ratio – Absolute zero (Velocity)</td>
</tr>
</tbody>
</table>

- Nominal – Categories (Colors)
- Ordinal – Ordered Categories (Sequence)
- Interval – Meaningful distances (Time)
- Ratio – Absolute zero (Velocity)

Stevens (1946, 1951)

Quantitative scales can correlate to Qualitative scales
- Example: Scores of 85% correct or greater get an “A”

Parametric statistics should not be used with qualitative scales.

**Statistical significance** should not be confused with **practical significance**
Example: Are you big enough to ride this roller coaster?

Theoretical Concepts: Safety, Park insurability, Liability

Measure: Physical Stature

Operational Definition: Height of child’s head against a vertical ruler.

ASSUMPTION: Height is the critical measure of stature.

ALL MEASURES UNDERGO INTERPRETATION

Measurement: Child stands next to a ruler

Measurement Scale: Distance (inches)

At amusement parks the scale is often just a cut-out clown figure and in this case the measure is the metric.

“If you are as tall as Puddles the Clown, you can ride this ride.”

Metric: Some minimum height. If that height is met, the child rides the ride. If not, the child does not ride the ride.
Crucial Point: Metrics come from Policy.

Metrics do not come from the underlying science, the theory, the theoretical concepts, the measures, the measurement methodology, the measurement scales, or any of that.

Policy: Do not kill the customer or get sued.

Metrics come from Policy. They do not magically spring from the measures or measurements.

Research sponsor is responsible for the policy.
Evaluating technologies for Cyber Protection Teams (CPTs)

**CPT Mission:** Defend priority DoD networks and systems against priority threats

**Performance goal:** *Detect, characterize, and mitigate before any damage can be done?*

Perhaps, however consider the following:

- **Decontextualization** - Mitigation might reveal your capabilities to the attacker. You may not always want to do that.
- **Reductive Thinking** - This proposed metric is a raw performance measure. It does not get at the “work system” level.

**Conclusion:** The measurement of cognitive work system performance must involve the application of multiple measures.
## CPT Task: Map a Cyberspace Network

### Critical Network Characteristics

- Number/Type of devices on network
- Applications/Services/Operating Systems
- Physical/Logical Architecture
- Communication paths
- High value systems (e.g., servers, system admin devices)
- Open ports
- Roles of Devices (e.g., web server, domain controller, user workstation)
- External connections
- Directory service information (e.g., Lightweight Directory Access Protocol (LDAP))
- User privileges and roles
- Software configurations
- Router configurations
- Normal (and aberrant) traffic patterns
Theoretical Concepts

• **Utility** – Does the tool help the team do the right things well?

• **Usability** – Does the tool work in the hands of real teams?

• **Acceptability** – Does the tool operate within the operational constraints of real teams?
# Measures and Metrics

<table>
<thead>
<tr>
<th>Theoretical Concept</th>
<th>Measure</th>
<th>Operational Definition</th>
<th>Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sufficiency</td>
<td>Number of tasks completed with tool</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Efficiency</td>
<td>Time required to complete assigned tasks</td>
<td>&lt;8 hours</td>
</tr>
<tr>
<td></td>
<td>Accuracy</td>
<td>Completeness and correctness of survey data</td>
<td>90% physical devices and paths enumerated</td>
</tr>
<tr>
<td></td>
<td>Data Integration</td>
<td>Types of data used to make map</td>
<td>PCAP, Config files, Netflow, SNMP, ICMP</td>
</tr>
<tr>
<td></td>
<td>Transparency</td>
<td>Ability to display what types of data were used</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Map Richness</td>
<td>Network attributes rendered on the map</td>
<td>All device types and physical routing</td>
</tr>
<tr>
<td></td>
<td>Exportability</td>
<td>Formats possible for exporting data and products</td>
<td>Visio, Image, and Data</td>
</tr>
<tr>
<td></td>
<td>User Feedback</td>
<td>Ease of use/learnability</td>
<td>60% positive</td>
</tr>
<tr>
<td></td>
<td>Map Interactivity</td>
<td>Ability to explore and annotate the map</td>
<td>Both</td>
</tr>
<tr>
<td></td>
<td>Support to Job</td>
<td>Prompts for normative processes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assistance Required</td>
<td>User requests for help</td>
<td>1/day</td>
</tr>
<tr>
<td></td>
<td>Network Load</td>
<td>Impact of network scans on the network</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>CPU Load</td>
<td>CPU usage over time</td>
<td>TBD</td>
</tr>
</tbody>
</table>

## Pitfalls

## Theory

## Example

## Summary
Measurement Challenges and Issues

Avoiding Decontextualization by Using Multiple Measures

Avoiding Reductive Thinking and Promoting Systems-level Thinking

Example: A new software support system improves performance on some task.

- Traditional Performance measurement would look only at HEAT measures: Hits, Errors, Accuracy, and Time
- This *faster-better-cheaper* techno-centric focus puts the worker in a “John Henry vs. the Steam Hammer” dilemma.
  - Worker feels like a slave to the machine.
  - Does the software tool promote continued learning and expertise?
  - Does it enhance worker intrinsic motivation?
Systems-level Measurement

- Cognitive work systems must be usable, useful, understandable and observable. Empirical evidence must accompany “deliverable.”

- Measures must support:
  - Evaluation of hypotheses concerning the nature of the cognitive work (e.g., synchronous versus asynchronous communication, effects of team experience, etc.)
  - Evaluation of the software tools themselves

- Methodology:
  - Study work
  - Operationally relevant tasks and conditions
  - Representative users
  - Include developers in assessments
  - Be prepared to be surprised
Summary

Theoretical Concepts → Measures → Operational Definitions → Measurements

BUT

Policy → Metrics

Metric - threshold for making a valuation or decision

“Universal Metrics” do not exist, because decisions are context sensitive

Contacts and References

LTC Stoney Trent, Ph.D.  satrent@cybercom.mil
Robert Hoffman, Ph.D.  rhoffman@ihmc.us

Selected References