Surveys in Test & Evaluation: The Good, The Bad, & The Ugly

Rebecca A. Grier, Ph.D.
Institute for Defense Analyses
DOT&E Guidance on Surveys

• Surveys are an important aspect of DOT&E evaluation of effectiveness and suitability

• Surveys are appropriate for quantitatively measuring operator and maintainer thoughts and opinions

• Use surveys only when appropriate

• Employ best practices for writing and administering surveys
  – Memo provides a best practices guide attachment
An Overview

What Will Be Covered

- What can’t be assessed by survey
- Some best practices for writing questions

What Won’t Be Covered

- Selecting academic surveys
- Formatting surveys
- Selecting response option type
- DOE with surveys
- Analysis of survey data
- Interviewing techniques
What is a Survey?

A systematic collection & analysis of data relating to the thoughts of a population.

What do you see?

Thoughts are relative; Requirements are not!

- Wants
- Thoughts
- Beliefs

Physical System Requirements
Examples Where Measurement by Survey Is Not Appropriate

• “The ship has protective clothing for every crew member.”
  – Count the protective clothing & compare to number of crew.

• “Engine exhaust levels in the mission bay do not exceed safety limits ...”
  – Measure exhaust levels with Portable Emissions Measurement System (PEMS) & compare to safety limits

• “Temperatures in primary work spaces were adequate.”
  – Measure by a thermometer & compare to requirements in MIL-STD-1472G
  – e.g., 5.5.2.1.4 part g: Limited thermal tolerance zones. Where hard physical work is to be required for more than two hours, an environment not exceeding... 25 °C (77 °F) shall be provided. ... shall be decreased 5.0 °C (9.0 °F) for complete chemical protective uniforms, 4.0 °C (7.0 °F) for intermediate clothing systems, and 3.0 °C (5.0 °F) for body armor.
A systematic collection & analysis of data relating to the thoughts of a population.

1. Collect specific data for a pre-defined purpose according to rules (i.e., not random thoughts).
2. The data collection rules determine the validity of the data & the statistical analyses possible.

<table>
<thead>
<tr>
<th>Who</th>
<th>Role</th>
<th>Sources of Error</th>
</tr>
</thead>
</table>
| Commissioner         | Defines Survey’s Purpose & Uses Information from Survey | • Not Enough Information  
|                      |                                                | • Wrong Information                                   |
| Participant/         | Gives Data                                     | • Answers Different Question  
| Respondent           |                                                | • Thinks Too Much                                     
|                      |                                                | • Doesn’t Think Enough                                 |
| Analyst              | Translates Data Into Information               | • Unable to analyze data  
|                      |                                                | • Data Aggregation                                     |
Question Writing Best Practices

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Item</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How many parts of a question are there?</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Knowledge Liability: Respondents Have Enough Information to Answer the Question

Singularity: Only 1 Idea Per Item

User Friendly: Items Do Not Require a Lot of Thought or Interpretation (e.g., short, clear, specific)

Neutrality: Items Do Not Imply Value Judgments
Items Are Not Emotionally Charged

Independence: Responses Will Not Affect Responses to Other Questions (Branching, Redundancy, etc…)
- Especially important when aggregating responses
- Item Response Theory & Internal Reliability

Available response options match items
All possible response options
1. The system is efficiently reliable.

User Friendly

Unclear what the goal of the question is. Therefore no recommended rewording.
2. Rate the overall ability of the system to provide situation awareness.
   Not Adequate    Adequate    DK/NA

2b. If not adequate, rate the degree to which this deficiency degrades effectiveness.
   Very Low        Low         Moderate       High       Very High
2. Rate the overall ability of the system to provide **situation awareness**. 

   Not Adequate    Adequate    DK/NA

2b. If not adequate, rate the degree to which this deficiency degrades **effectiveness**. 

   Very Low    Low    Moderate    High    Very High

Knowledge Liability
2. Rate the overall ability of the system to provide **situation awareness**.

   Not Adequate    Adequate    DK/NA

2b. **If not adequate**, rate the degree to which this deficiency degrades **effectiveness**.

   Very Low    Low    Moderate    High    Very High

   Knowledge Liability

   Independence
2. Rate the overall ability of the system to provide situation awareness.
   Not Adequate    Adequate    DK/NA

2b. If not adequate, rate the degree to which this deficiency degrades effectiveness.
   Very Low    Low    Moderate    High    Very High

   Knowledge Liability
   Independence
   All Possible Responses

6. I trust the information provided by the system.

   Strongly Agree    Somewhat Agree    Slightly Agree    Slightly Agree    Somewhat Disagree    Strongly Disagree
3. Is the training materials complete?
3. Is the training materials complete?

Knowledge Liability

User Friendly – grammar

5. I felt as if I needed more training.
4. Rate the acceptability of system’s Launch Acceptability Region (LAR) displays provided to support accurate and timely system employment.
4. Rate the acceptability of system’s Launch Acceptability Region (LAR) displays provided to support **accurate** and **timely** system employment.

Knowledge Liability
4. Rate the acceptability of system’s Launch Acceptability Region (LAR) displays provided to support **accurate and timely** system employment.

Knowledge Liability

Singularity
4. Rate the acceptability of system’s Launch Acceptability Region (LAR) displays provided to support accurate and timely system employment.

Knowledge Liability
Singularity
User Friendly
(repeated and unnecessary words)

3. The Launch Acceptability Region (LAR) displays are helpful.
5. The SSO functionality increases my productive time within the clinic by reducing the amount of time I spend logging into different applications when documenting the healthcare provided.

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Somewhat Agree</th>
<th>Slightly Agree</th>
<th>Slightly Agree</th>
<th>Somewhat Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
</table>
28 July 2014
5. The SSO functionality increases my productive time within the clinic by reducing the amount of time I spend logging into different applications when documenting the healthcare provided.

Neutrality
User Friendly – 27 words!

4. The SSO function is useful.
6. Rate the overall usefulness of the report: **Accuracy**

<table>
<thead>
<tr>
<th>Completely unacceptable</th>
<th>Largely unacceptable</th>
<th>Somewhat unacceptable</th>
<th>Somewhat acceptable</th>
<th>Largely acceptable</th>
<th>Completely acceptable</th>
</tr>
</thead>
</table>

28 July 2014
6. Rate the overall **usefulness** of the report:

**Accuracy**

<table>
<thead>
<tr>
<th>Completely unacceptable</th>
<th>Largely unacceptable</th>
<th>Somewhat unacceptable</th>
<th>Somewhat acceptable</th>
<th>Largely acceptable</th>
<th>Completely acceptable</th>
</tr>
</thead>
</table>

**Knowledge Liability**

Response Options Don’t Match Item

**User Friendly**

(What is being rated: report acceptability, usefulness, or accuracy?)

5. The content of the report is useful.

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Somewhat Agree</th>
<th>Slightly Agree</th>
<th>Slightly Agree</th>
<th>Somewhat Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
</table>
7. Based on your responses above, rate the acceptability of the system.
7. Based on your responses above, rate the acceptability of the system.

Independence
User Friendly
(It is the analyst’s job to rate the acceptability of the system)

8. I would like to use this system to accomplish the mission.
Some Words to Avoid In Writing Surveys

- Accurate
- Timely
- Situation Awareness
- Effective
- Efficient
- n/a
- And
- Each
- All
- Never
- None
- Better
- Easier
- Improved
- Based on
- If
- Considering
Questions?
2. Rate the adequacy of air-search radar & combat system to correctly decide to engage/not engage each track per Combat System Engagement Doctrine.
2. Rate the adequacy of air-search radar & combat system to **correctly** decide to engage/not engage each track per **Combat System Engagement Doctrine**.

Knowledge Liability
2. Rate the adequacy of air-search radar & combat system to correctly decide to engage/not engage each track per Combat System Engagement Doctrine.

Knowledge Liability

Singularity

User Friendly – 22 words!

2. I trusted the system’s engagement decisions.
7. The amount and type of training provided to the X position allowed them to employ the system.
7. The amount and type of training provided to the X position allowed them to employ the system.

Knowledge Liability
7. The amount and type of training provided to the X position allowed them to employ the system.

Knowledge Liability

Singularity

7. I felt as if I needed more training.