

---

# Science of Survey Design: Overview of DOT&E Guidance

Catherine Warner





# Overview

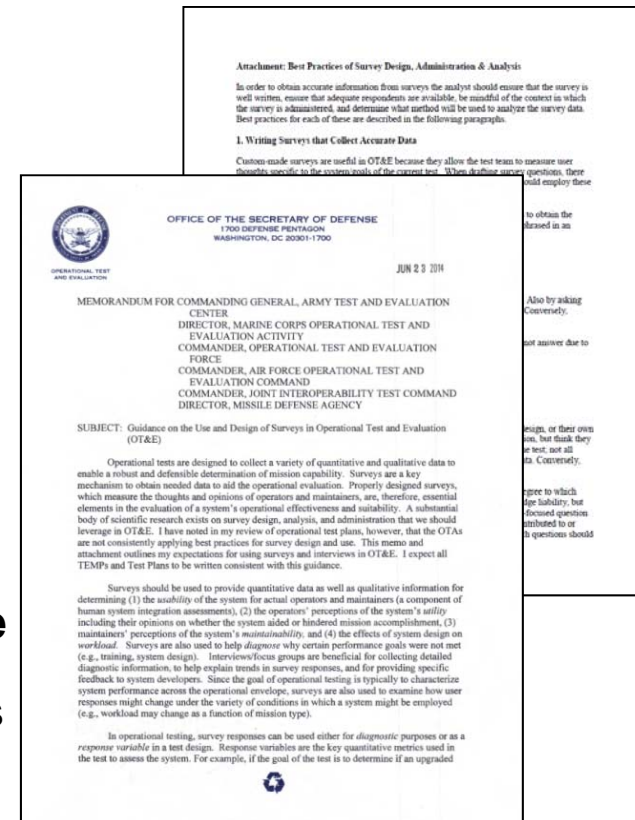
---

- **Review key elements of DOT&E guidance on the use of surveys and focus groups in operational testing**
- **Example applications of the guidance**



# Surveys in OT&E (June 2014): High Level guidance

- Surveys measure the thoughts and opinions of operators and maintainers
- Surveys are an essential element of OT&E
  - Effectiveness
    - » User assessment of system's aid to mission
    - » User assessment of system utility
    - » Diagnostic and contextual information
  - Suitability
    - » Maintainability
    - » Safety
    - » Human Factors
    - » Usability
    - » Workload
    - » Training
- Surveys should not be used to measure performance or as an absolute measure of situational awareness
  - Surveys can be used to identify problems in these areas
- There should always be a clear intent behind survey construction (this is documented in the test plan)



*Surveys are essential for measuring thoughts*



# Surveys in OT&E (June 2014): High Level guidance

- **It is essential to understand the goal of why you are conducting the survey**
  - Emphasized in each recommendation made in the memorandum
- **There is no one-size-fits all solution**
- **The goals of your test and your survey will influence the design decisions that you make**
  - Surveys versus interviews/focus groups
  - Quantitative versus qualitative
  - Response variable versus diagnostic measure
  - Empirically vetted versus custom-made
- **In some cases, asking the operator may be the only cost effective data collection method available for certain measures**
  - In these cases, data collection sheets should be developed and clearly differentiated from surveys
  - Data collection sheets for operators/maintainers should be limited to prevent “survey fatigue”

#### Attachment: Best Practice of Survey Design, Administration & Analysis

In order to obtain accurate information from surveys the analyst should ensure that the survey is well written, ensure that adequate respondents are available, be mindful of the context in which the survey is administered, and determine what method will be used to analyze the survey data. Best practices for each of these are described in the following paragraphs.

#### 1. Writing Surveys that Collect Accurate Data

Custom-made surveys are useful in OT&E because they allow the test team to measure user thoughts specific to the system goals of the current test. When drafting survey questions, there could employ these

OFFICE OF THE SECRETARY OF DEFENSE  
7000 DEFENSE PENTAGON  
WASHINGTON, DC 20301-1700

JUN 23 2014

MEMORANDUM FOR COMMANDING GENERAL, ARMY TEST AND EVALUATION CENTER  
DIRECTOR, MARINE CORPS OPERATIONAL TEST AND EVALUATION ACTIVITY  
COMMANDER, OPERATIONAL TEST AND EVALUATION FORCE  
COMMANDER, AIR FORCE OPERATIONAL TEST AND EVALUATION COMMAND  
COMMANDER, JOINT INTEROPERABILITY TEST COMMAND  
DIRECTOR, MISSILE DEFENSE AGENCY

SUBJECT: Guidance on the Use and Design of Surveys in Operational Test and Evaluation (OT&E)

Operational tests are designed to collect a variety of quantitative and qualitative data to enable a robust and defensible determination of mission capability. Surveys are a key mechanism to obtain needed data to aid the operational evaluation. Properly designed surveys, which measure the thoughts and opinions of operators and maintainers, are, therefore, essential elements in the evaluation of a system's operational effectiveness and suitability. A substantial body of scientific research exists on survey design, analysis, and administration that we should leverage in OT&E. I have noted in my review of operational test plans, however, that the OTAs are not consistently applying best practices for survey design and use. This memo and attachment outlines my expectations for using surveys and interviews in OT&E. I expect all TEMPs and Test Plans to be written consistent with this guidance.

Surveys should be used to provide quantitative data as well as qualitative information for determining (1) the usability of the system for actual operators and maintainers (a component of human system integration assessments), (2) the operators' perceptions of the system's *usability* including their opinions on whether the system aided or hindered mission accomplishment, (3) maintainers' perceptions of the system's *maintainability*, and (4) the effects of system design on *workload*. Surveys are also used to help *diagnose* why certain performance goals were not met (e.g., training, system design). Interviews/focus groups are beneficial for collecting detailed diagnostic information, to help explain trends in survey responses, and for providing specific feedback to system developers. Since the goal of operational testing is typically to characterize system performance across the operational envelope, surveys are also used to examine how user responses might change under the variety of conditions in which a system might be employed (e.g., workload may change as a function of mission type).

In operational testing, survey responses can be used either for *diagnostic* purposes or as a *response variable* in a test design. Response variables are the key quantitative metrics used in the test to assess the system. For example, if the goal of the test is to determine if an upgraded

to obtain the desired in in

Also by asking Conversely,

not answer due to

design, or their own ion, but think they as test, not all its. Conversely,

agree to which figure liability, but focused question attributed to or fi questions should

**Overarching theme: Understand the goal of surveys**



# Surveys in OT&E (June 2014): Best practices

- **Question Development**

- Improve accuracy of data collected by employing well known rules:
  - » Ensure neutrality
  - » Avoid knowledge liabilities
  - » Ensure questions are user friendly
  - » Ensure singularity
  - » Ensure independence
- Minimize survey length
  - » Many surveys are too long

- **Administration**

- Well-designed surveys can be compromised by a poor administration process
- Preserve neutrality of the survey
- Ensure respondent confidentiality
- Be administered in a timely fashion
- Consistent administration procedures

Attachment: Best Practices of Survey Design, Administration & Analysis

In order to obtain accurate information from surveys the analyst should ensure that the survey is well written, ensure that adequate respondents are available, be mindful of the context in which the survey is administered, and determine what method will be used to analyze the survey data. Best practices for each of these are described in the following paragraphs.

#### 1. Writing Surveys that Collect Accurate Data

Custom-made surveys are useful in OT&E because they allow the test team to measure user thoughts specific to the system goals of the current test. When drafting survey questions, there should employ these

OFFICE OF THE SECRETARY OF DEFENSE  
1700 DEFENSE PENTAGON  
WASHINGTON, DC 20301-1700

JUN 23 2014

MEMORANDUM FOR COMMANDING GENERAL, ARMY TEST AND EVALUATION CENTER  
DIRECTOR, MARINE CORPS OPERATIONAL TEST AND EVALUATION ACTIVITY  
COMMANDER, OPERATIONAL TEST AND EVALUATION FORCE  
COMMANDER, AIR FORCE OPERATIONAL TEST AND EVALUATION COMMAND  
COMMANDER, JOINT INTEROPERABILITY TEST COMMAND  
DIRECTOR, MISSILE DEFENSE AGENCY

SUBJECT: Guidance on the Use and Design of Surveys in Operational Test and Evaluation (OT&E)

Operational tests are designed to collect a variety of quantitative and qualitative data to enable a robust and defensible determination of mission capability. Surveys are a key mechanism to obtain needed data to aid the operational evaluation. Properly designed surveys, which measure the thoughts and opinions of operators and maintainers, are, therefore, essential elements in the evaluation of a system's operational effectiveness and suitability. A substantial body of scientific research exists on survey design, analysis, and administration that we should leverage in OT&E. I have noted in my review of operational test plans, however, that the OTAs are not consistently applying best practices for survey design and use. This memo and attachment outlines my expectations for using surveys and interviews in OT&E. I expect all TEMPs and Test Plans to be written consistent with this guidance.

Surveys should be used to provide quantitative data as well as qualitative information for determining (1) the usability of the system for actual operators and maintainers (a component of human system integration assessments), (2) the operators' perceptions of the system's utility, including their opinions on whether the system aided or hindered mission accomplishment, (3) maintainers' perceptions of the system's maintainability, and (4) the effects of system design on workload. Surveys are also used to help diagnose why certain performance goals were not met (e.g., training, system design). Interview/focus groups are beneficial for collecting detailed diagnostic information, to help explain trends in survey responses, and for providing specific feedback to system developers. Since the goal of operational testing is typically to characterize system performance across the operational envelope, surveys are also used to examine how user responses might change under the variety of conditions in which a system might be employed (e.g., workload may change as a function of mission type).

In operational testing, survey responses can be used either for diagnostic purposes or as a response variable in a test design. Response variables are the key quantitative metrics used in the test to assess the system. For example, if the goal of the test is to determine if an upgraded




# Focus Groups & System Utility February 2015

- Surveys & Focus Groups serve different purposes in T&E
- Focus Groups do not collect statistically analyzable data
  - Conformity
  - Polarization
- Focus Groups support interpretation of survey responses and observed performance.

• Do you have any additional comments about the system? [open response]

ake this why you system said the For those to be he use and



OFFICE OF THE SECRETARY OF DEFENSE  
1700 DEFENSE PENTAGON  
WASHINGTON, DC 20301-1700

FEB 24 2015

MEMORANDUM FOR COMMANDER, ARMY TEST AND EVALUATION COMMAND

SUBJECT: Discussion on the Use and Design of Surveys


Thank you for your 22 December 2014 response to my recently issued Guidance on the design and use of surveys in Operational Test and Evaluation (OT&E) (memorandum dated 23 June 2014). I am pleased to see that you are taking actions to put this new guidance into practice, and training your staff on these methods and principles to ensure the best data are collected from surveys, interviews, and focus groups. Your efforts will improve our data collection particularly for surveys and interviews which can play a significant role in aiding the determination of effectiveness and/or suitability. As I have stated before, I also will make my staff available to aid in your training efforts, if you so desire.

**Focus Groups**

Given that much discussion has occurred over the proper lines of questioning used for focus group, I would like to bring clarity and direction about the role of and best practices for conducting focus groups. Focus groups have become a significant component of Army test and evaluation, particularly during Network Integration Evaluations (NIEs).<sup>1</sup> I view focus groups as useful, often essential, venues to elicit operator opinions; however, focus groups should not be the sole source of operator opinion data. Focus groups, group interviews, and the like, can easily be affected by group dynamics. These dynamics can take on several forms, including *conformity*, where members of the group might change their opinion to conform to the group's emerging opinion or one particular person's (e.g., commanding officer) opinion, and *group polarization*, where group members tend to form opinions that are more extreme than individual opinions. Because of these effects, it is difficult to obtain an accurate measure of individuals' opinions because individual opinions can become systematically biased as a direct result of the group setting.<sup>2</sup> Any quantitative data obtained from focus groups are necessarily dependent, meaning that if you have ten participants in your focus group, you do not necessarily have ten independent observations. In fact, it may be the case that you only have one unique observation. Consequently, the use of traditional statistical tests (t-tests, chi-square, tests of proportions, etc.)

<sup>1</sup> As a point of clarification, the principles I discuss here also apply to after action reviews, hot-washes, and mission debriefings, since the expression of operator opinions is likely to occur during these periods as well.

<sup>2</sup> There have been many experiments in psychology investigating conformity and group pressure. Some references on this area of study include A. Asch, "The role of discussion in changing opinion regarding a matter of fact," *The Journal of Abnormal and Social Psychology* (1952), H. C. Kelman, "Compliance, identification, and internalization: Three processes of attitude change," *Journal of conflict resolution* (1958), P. Kidd and M. Parshall, "Oetting the Focus and the Group: Enhancing Analytical Rigor in Focus Group Research," *Qualitative Health Research* (2000), and S. Moscovici and M. Zavalloni, "The Group as a Polarizer of Attitudes," *Journal of Personality and Social Psychology*, Vol. 12, No. (1969). This list is not exhaustive; other references can be provided.






# Focus Groups & System Utility

## February 2015

- **Opinions of operators & maintainers of system utility is important to evaluation**
  - Survey question - Likert
  - I would take this system to war.
  - I would like to use this system to accomplish the mission.
- **Utility is not a measure of system effectiveness**
  - Perception of utility by user
- **Utility & usability are interdependent**
- **Utility may be affected by training**

• Do you have any additional comments about the system? [open response]

ake this  
why you  
system  
said the  
For those  
to be  
he use and



OFFICE OF THE SECRETARY OF DEFENSE  
1700 DEFENSE PENTAGON  
WASHINGTON, DC 20301-1700

FEB 24 2015

MEMORANDUM FOR COMMANDER, ARMY TEST AND EVALUATION COMMAND

SUBJECT: Discussion on the Use and Design of Surveys


Thank you for your 22 December 2014 response to my recently issued Guidance on the design and use of surveys in Operational Test and Evaluation (OT&E) (memorandum dated 23 June 2014). I am pleased to see that you are taking actions to put this new guidance into practice, and training your staff on these methods and principles to ensure the best data are collected from surveys, interviews, and focus groups. Your efforts will improve our data collection particularly for surveys and interviews which can play a significant role in aiding the determination of effectiveness and/or suitability. As I have stated before, I also will make my staff available to aid in your training efforts, if you so desire.

**Focus Groups**

Given that much discussion has occurred over the proper lines of questioning used for focus group, I would like to bring clarity and direction about the role of and best practices for conducting focus groups. Focus groups have become a significant component of Army test and evaluation, particularly during Network Integration Evaluations (NIEs).<sup>1</sup> I view focus groups as useful, often essential, venues to elicit operator opinions; however, focus groups should not be the sole source of operator opinion data. Focus groups, group interviews, and the like, can easily be affected by group dynamics. These dynamics can take on several forms, including *conformity*, where members of the group might change their opinion to conform to the group's emerging opinion or one particular person's (e.g., commanding officer) opinion, and *group polarization*, where group members tend to form opinions that are more extreme than individual opinions. Because of these effects, it is difficult to obtain an accurate measure of individuals' opinions because individual opinions can become systematically biased as a direct result of the group setting.<sup>2</sup> Any quantitative data obtained from focus groups are necessarily dependent, meaning that if you have ten participants in your focus group, you do not necessarily have ten independent observations. In fact, it may be the case that you only have one unique observation. Consequently, the use of traditional statistical tests (t-tests, chi-square, tests of proportions, etc.)

<sup>1</sup> As a point of clarification, the principles I discuss here also apply to after action reviews, hot-washes, and mission debriefings, since the expression of operator opinions is likely to occur during these periods as well.

<sup>2</sup> There have been many experiments in psychology investigating conformity and group pressure. Some references on this area of study include A. Asch, "The role of discussion in changing opinion regarding a matter of fact," *The Journal of Abnormal and Social Psychology* (1952), H. C. Kelman, "Compliance, identification, and internalization: Three processes of attitude change," *Journal of conflict resolution* (1958), P. Kidd and M. Parshall, "Oetting the Focus and the Group: Enhancing Analytical Rigor in Focus Group Research," *Qualitative Health Research* (2000), and S. Moscovici and M. Zavalloni, "The Group as a Polarizer of Attitudes," *Journal of Personality and Social Psychology*, Vol. 12, No. (1969). This list is not exhaustive; other references can be provided.





# Neutral Responses

## March 2015

- Neutral responses should be used appropriately
- When used neutral responses should be mid point in a response continuum
- Not applicable options should be avoided
- Parametric analyses should be utilized when assumptions are met

6. In terms of the analysis, we have advocated for using parametric models that treat Likert data as continuous (aka regression) when appropriate. We should always  
that this is  
in the first



OFFICE OF THE SECRETARY OF DEFENSE  
1700 DEFENSE PENTAGON  
WASHINGTON, DC 20301-1700

APR 02 2015

MEMORANDUM FOR ASSISTANT DEPUTY UNDER SECRETARY OF THE ARMY FOR  
TEST AND EVALUATION

SUBJECT: Discussion on Including Neutral Responses on Survey Questions

In your March 27, 2015 email to me, you suggested that all surveys given in operational testing should include a "don't care" option rather than forcing a choice by having an even number of response choices, symmetrical across positive and negative opinions. As I have stated in two memoranda dated June 23, 2014 and February 24, 2015 on the Use and Design of Surveys in Operational Test and Evaluation, all surveys should adhere to best practices that have been established by the academic community, and my answer to your question is consistent with that guidance.

Neutral responses should be used only when appropriate. Unfortunately, neutral responses have been overused in past operational testing. In general, neutral responses should be avoided when it is clear that the operator will have some opinion. For example, if the survey is asking whether "This system is easy to operate," the operator should have an opinion, and consequently, a neutral response is not appropriate. On the other hand, neutral responses can be interesting and should be included in cases where a direct comparison is being made. For example, if the survey is asking whether "The new system was easier to use than the legacy system," a neutral response is interesting and the operators may have no preference.

The number of times that neutral responses are needed can be minimized by tailoring the surveys. We should avoid asking questions that do not apply to the some or all of the population. For example, testers should avoid giving the same survey to both operators and maintainers with the expectation that some of the questions will apply to only one of the groups. In order to maintain high motivation for completing a survey, it is important that the survey be tailored (e.g., two different surveys for operators and maintainers) to minimize the number of questions each person must answer and limit the possibility of asking irrelevant questions.

If a neutral response is appropriate, it should always be asked in the form of a 5 or 7 point continuum response scale, with the neutral response providing the midpoint. Providing a neutral response or "no opinion" outside of the scale should always be avoided. In those cases, it is common practice to treat the no opinion option as missing data, which provide no insight for our analyses. Furthermore, because of the small sample sizes we have in operational testing, we need the increased power that comes from forcing the respondent to provide the direction they are leaning toward. Research has shown that increasing the number of response options improves both reliability and validity of the question, but with diminishing returns after 7 point







# Surveys in TEMP's & Test Plans

---

- **TEMP should indicate which COIs will be assessed using surveys, focus groups, or interviews**
- **Test Plans should contain the following information**
  - Specific survey
    - » Empirically vetted
    - » Custom-made
  - When will the survey be administered
    - » At break points, end of test, event driven?
  - The goal of the survey
    - » Why is it important to collect these data at these points in the test?
  - How the survey will be administered
    - » Verbally, electronically, paper
  - Who will complete the survey
    - » Which users/maintainers in the test?
  - How will data be vetted, stored, & analyzed



## DOT&E Vetted Example Questions

---

- I would take this system to war.
- I would like to use this system to accomplish the mission.
- The instructor presented the material clearly.
- I feel as though additional training is needed.
- The \_(e.g., work station, cockpit)\_ is well organized.
- I did not have the information needed to \_\_ (e.g., execute the mission, perform a specific task) \_\_.
- It was difficult to \_ (e.g., perform a specific task) \_.
- \_(e.g., Equipment, Controls, Information, Features, Applications)\_ are easily accessible.
- Are there any improvements that you would make to the system?
- Please comment on any safety concerns that you have.
- Many others...

*Some of the Questions that meet the best practices*



# GBS Survey Results

- **Survey ratings are generally positive, but technical orders were insufficient to configure the legacy receive suite for the GBS DECC broadcast, requiring GBSOC help**
  - 7-point scale used with rating levels 1-3 defined as unacceptable and 4-7 as sufficient or better

Rating Topic	Number of Respondents	Mode	Number of Answers $\leq 3$	
Receive suite operator ratings of overall acceptability of GBS DECC products	4	5	0	
Receive Suite operator ratings of the overall acceptability of the process to develop a GMR	6	7	0	
GBSOC operator ratings of the GBS DECC for planning broadcast missions	12	6	0	
Receive suite operator ratings of the GBSOC Help Desk function	18	6	1	
GBSOC operator ratings of GBS DECC safety	16	7	0	
GBSOC and receive suite operator ratings of GBS DECC documentation	18	6	1	
GBSOC operator ratings of training	14	6	0	
GBSOC operator ratings of human factors topics	13	6	0	
GBSOC operator ratings of the overall acceptability of the GBS DECC planning software	16	6	0	

Acronyms this slide: Defense Enterprise Computing Center (DECC); Global Broadcast Service (GBS); GBS Operations Center (GBSOC)



# Remainder of Today

---

- **Introduction to Surveys**
- **Human Measurement**
- **Selecting Empirically Vetted Surveys**
- **Custom-Made Surveys**
- **ABIS Case Study**
- **Administration & Analysis**
- **Air Force DCGS Case Study**