



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

JAN 06 2017

OPERATIONAL TEST  
AND EVALUATION

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

SUBJECT: Survey Pre-Testing and Administration in Operational Test and Evaluation

- References: (a) Babbitt, B. A., & Nystrom, C. O. (1989). *Questionnaire Construction Manual*. U.S. Army Research Institute for the Behavioral and Social Sciences.
- (b) Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, phone, mail, and mixed-mode surveys: the tailored design method*. John Wiley & Sons.
- (c) Fink, A. & Kosecoff, J. (2015). *How to conduct surveys: A step-by-step guide*. Sage Publications.
- (d) Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology*. John Wiley & Sons.

This memorandum builds on existing guidance to clarify pretesting and administration procedures for surveys and articulates my expectations for the type of survey-related information that Operational Test Agencies (OTAs) should include in Test and Evaluation Master Plans (TEMPs) and Test Plans. Previously, I issued two memoranda on survey methods in operational test and evaluation dated June 23, 2014 and April 2, 2015. This guidance addressed best practices in survey design, highlighting the need for well-written questions and also encouraged the use of empirically-vetted instruments, such as the System Usability Scale (SUS) and National Aeronautics and Space Administration Task Load Index (NASA-TLX). The test and evaluation (T&E) community responded by promptly implementing this guidance, and the quality of survey instruments has improved significantly during the past two years. However, we continue to identify areas where we can improve rigor and the quality of data obtained, especially if survey data are the only means to evaluate specific measures in a test design. Survey data quality, and therefore its usefulness in my evaluation, is limited by poor pre-testing and administration procedures.

The following sections address two major issues. The first involves pretesting custom surveys developed for a specific test. The second involves administering surveys and is applicable to both custom and empirically-vetted surveys. The information below supplements my earlier guidance on the use of surveys.



## Pre-Testing Survey Instruments

Pretesting surveys (often referred to as pilot-testing) is a deliberate review of the survey to ensure that respondent answers will be what the testers need and are useful for the required analyses. Pretesting surveys should not be confused with the traditional pilot test – an event that immediately precedes the operational test to confirm that all data collection procedures are working properly. Pretesting should occur as part of the test planning process, prior to submitting the Operational Test Plan for DOT&E approval.

Pretesting is widely and strongly recommended by survey experts in academia, industry, and the military. The Army Research Institute's (ARI's) *Questionnaire Construction Manual* presents the most straightforward mandate for pretesting [emphasis in original]:

Pretesting is an important and essential procedure to follow before administering any questionnaire... Pretesting may seem to some uninformed individuals to be a waste of time, especially when the author may have asked several people in his/her office to critique the questions, or perhaps even asked a questionnaire specialist to critique it. However, pretesting is an investment that is well worthwhile. It is crucial if the decision that will result from the questionnaire is of any importance.

Chapter XI of ARI's *Questionnaire Construction Manual* goes on to describe the method they recommend for pretesting a survey. Other resources are available that describe how to execute a pretest.<sup>1</sup> One of these sources, or another reputable source, should be identified in the TEMP or supporting documentation to describe how the survey will be pretested. The different methods of pretesting described in these sources share some common attributes. First, pretesting requires an outside view – from people who are not on the test team or were not involved with creating the survey. Second, pretesting requires either representative users or Subject Matter Experts (SMEs) on the topic being surveyed. While there are various methods of pretesting, they provide different kinds of information and using several different methods is important for a comprehensive test. In an operational test, at least one method of pretesting must be used. A custom survey addressing a critical question, such as one used to evaluate a specific measure, should be pretested using multiple methods.

Developing a custom survey is a significant investment of time and resources. Development timeline and resources required for pretesting the custom survey, such as outside experts or representative users, should be documented in the TEMP or supporting documentation. This documentation should identify:

1. Operational testing that will use custom survey questions
2. When the custom surveys will be developed
3. How the surveys will be pretested
4. When the revised surveys will be completed, including staffing timelines
5. Resources, such as experts or representative users, required for pretesting

---

<sup>1</sup> Chapter 6 of Dillman's *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*, Chapter 8 of Groves' *Survey Methodology*, and Chapter 3 of Fink's *How To Conduct Surveys*.

Along with the Operational Test Plan, the revised, pretested custom surveys should be submitted no later than 60 days prior to the test date for my review. All surveys, even those crafted by teams of experts, should expect to find faults during pretesting that need correction. Therefore, OTAs should allow adequate time for pretesting to adjust surveys to meet test plan approval deadlines.

## **Survey Administration Plans**

How surveys are administered during operational tests affects the kinds of conclusions that testers can draw from the data. Surveys that are administered systematically across test conditions produce insights into how human-system interactions change under different operational conditions and the impact of those interactions on system performance. Such data are critical for evaluating the effectiveness and suitability of military systems – especially those measures that only have surveys as a data source. Unfortunately, systematic administration is often the exception in operational testing rather than the rule. I routinely see test plans that do not describe the test conditions under which surveys will be administered or which users they will administer surveys to. Without this information, survey data analyses become less meaningful, if not impossible. Furthermore, submitting test plans without administration plans or with administration plans that lack detail restricts DOT&E's ability to provide feedback to improve the likelihood that the survey data collected during operational tests provides insights into the effectiveness and suitability of human-system interactions.

Administering surveys systematically requires careful planning. Testers must be specific about the goals of each survey measure. For example, simply stating that you will measure operator workload is not sufficient. Rather, testers should identify whether they will measure changes in workload across test conditions or if they plan to evaluate how workload affects system performance and be able to justify how specific survey questions will help them accomplish their evaluation goals. Testers must then identify test design and demographic factors that are relevant to the goals of the measure and plan to administer surveys across those test conditions. In addition, testers should report how users will be spread across test conditions, as this affects the kind of analyses that are possible. In particular, they should discuss whether users will undergo all test conditions (within-subjects design), if a unique group of users will undergo each test condition (between-subjects design), or if users will undergo some test conditions, but not others (mixed design). Distributing users across the test space systematically can increase the statistical power of survey data and reduce the number of times surveys need to be administered during operational tests.

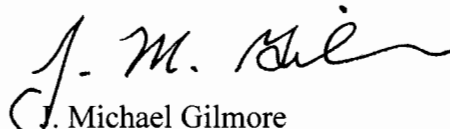
The Operational Test Plan (or accompanying documentation) should include a detailed survey administration plan discussing these issues along with the final survey instruments no later than 60 days prior to the test date for my review. The survey administration plan should include:

1. The goal of each survey measure, including how it will be analyzed
2. The number of users in the test
3. How (or if) users will be separated into groups

4. The test conditions that each group will undergo
5. When surveys will be administered throughout the test
6. How survey data from users will be mapped to test conditions, performance measures, and demographic data

A sample administration plan is provided in the Attachment for reference.

I appreciate the efforts the OTAs have made so far in improving survey usage in operational test and evaluation (OT&E). By pre-testing survey instruments and developing quality administration plans in the future we will ensure that the valuable input of our operators and maintainers is accurately represented in our OT&E Reports. My point of contact on this matter is Dr. Catherine Warner. She can be reached at (703) 697-3655.

  
J. Michael Gilmore  
Director

Attachment:  
As stated

# Attachment

## Example Survey Administration Plan

*This is a notional example of a survey administration plan for a military system.*

### Modular Cargo Vehicle System (MCVS)

The MCVS consists of 2.5-ton variant and 5-ton variant that support the Army across a range of military operations, but primarily transport cargo and troops in all weather and terrain conditions. The notional test design for MCVS is provided in Figure 1. The primary response variable of interest for the system is the average convoy speed across different types of terrain. As you can see, average convoy speed will be assessed across 5 factors: load (full vs. empty), light condition (day vs. night), system (MCVS vs. legacy), variant (2.5-ton vs. 5.0-ton), and trailer (with vs. without).

|        |         |                 | Full Load |       | Empty Load |       |
|--------|---------|-----------------|-----------|-------|------------|-------|
|        |         |                 | Day       | Night | Day        | Night |
| Legacy | 2.5-ton | With trailer    | 1         | 2     | 2          | 1     |
|        |         | Without trailer | 2         | 1     | 1          | 2     |
|        | 5.0-ton | With trailer    | 2         | 1     | 1          | 2     |
|        |         | Without trailer | 1         | 2     | 2          | 1     |
| MCVS   | 2.5-ton | With trailer    | 2         | 1     | 2          | 1     |
|        |         | Without trailer | 2         | 1     | 1          | 2     |
|        | 5.0-ton | With trailer    | 1         | 2     | 1          | 2     |
|        |         | Without trailer | 1         | 2     | 2          | 1     |

← Number of missions completed under each set of test conditions.

**Figure 1. Notional MCVS Test Design**

The survey administration plan will be integrated into the larger test design for the system. Imagine that in addition to the average convoy speed, the OTA is also interested in measuring operators' comfort and workload under various conditions. For instance, they may be interested in understanding whether MCVS is more comfortable and produces lower workload than the legacy system under different light conditions, loads, and in the presence or absence of a trailer. However, they are not interested in measuring differences in comfort and workload by variant (2.5-ton vs. 5.0-ton) because the design of the variants is very similar. The administration plan should indicate which subset of factors the OTA believes will affect operator comfort and workload. The factor selection process is depicted in Figure 2.

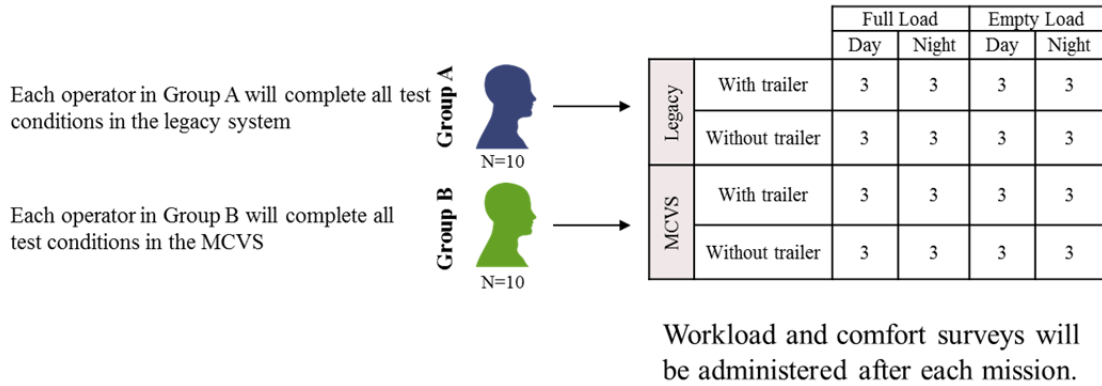
|        |         |                 | Full Load |       | Empty Load |       |
|--------|---------|-----------------|-----------|-------|------------|-------|
|        |         |                 | Day       | Night | Day        | Night |
| Legacy | 2.5-ton | With trailer    | 1         | 2     | 2          | 1     |
|        |         | Without trailer | 2         | 1     | 1          | 2     |
|        | 5.0-ton | With trailer    | 2         | 1     | 1          | 2     |
|        |         | Without trailer | 1         | 2     | 2          | 1     |
| MCVS   | 2.5-ton | With trailer    | 2         | 1     | 2          | 1     |
|        |         | Without trailer | 2         | 1     | 1          | 2     |
|        | 5.0-ton | With trailer    | 1         | 2     | 1          | 2     |
|        |         | Without trailer | 1         | 2     | 2          | 1     |

Subset →

|          |                 | Full Load |       | Empty Load |       |
|----------|-----------------|-----------|-------|------------|-------|
|          |                 | Day       | Night | Day        | Night |
| Legacy   | With trailer    | 3         | 3     | 3          | 3     |
|          | Without trailer | 3         | 3     | 3          | 3     |
| Upgraded | With trailer    | 3         | 3     | 3          | 3     |
|          | Without trailer | 3         | 3     | 3          | 3     |

**Figure 2. Factor Selection Process for the Survey Administration Plan**

Additionally, the OTA should indicate the number of users, how they will be spread across the test conditions, and when surveys will be administered. This can be provided in the text, a figure, or any combination of text and figure needed to clearly depict the administration process. An example is provided in Figure 3.



**Figure 3. Survey Administration Plan**

In this example, there are 20 users half of which will complete each test condition using the legacy system and the other half will complete each test condition in the MCVS – these are referred to as Group A and Group B, respectively. Each operator will complete 3 missions under each set of test conditions. Workload and comfort surveys will be administered after each mission.