FEB 2 4 2015

OPERATIONAL TEST
AND EVALUATION

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).[1] 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.[2] 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.)

---

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

[2] There have been many experiments in psychology investigating conformity and group pressure. Some references on this area of study include A. Jenness, "The role of discussion in changing opinion regarding a matter of fact." *The Journal of Abnormal and Social Psychology* (1932), H. C. Kelman, "Compliance, identification, and internalization: Three processes of attitude change." *Journal of conflict resolution* (1958), P. Kidd and M. Parshall, "Getting 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.

is not possible without significant design changes.[3] While we have seen cases where the group dynamic has apparently not affected individuals' responses, we must be aware that the potential exists for skewed data to be collected.[4] Furthermore, even if no group dynamic *appears* to be occurring, it is impossible to verify that individual responses are unbiased. Because of this danger, focus groups should, therefore, be used primarily as diagnostic tools to further understand the data from the written surveys, and elicit discussion from the users on why their survey responses might have been particularly negative or positive, or diagnose a cause for a mixed response across different demographics or other factors in the test design.

Hence, I agree with concerns previously expressed to my staff, that focus groups should not elicit a polled binary (yes/no) response from the users to obtain accurate survey data. As my previous guidance stated, binary responses are discouraged since humans will not all agree on the dividing line between positive and negative responses, and furthermore, the data are limited to binomial statistics, providing poor statistical precision. Combining these problems with the potentially correlated or skewed count data obtained in group settings, argues for eliminating this practice from NIEs or other venues.

Using the focus groups in a diagnostic role will require the written surveys to be administered first. The test team should, therefore, examine the survey responses immediately after administering them to look for trends (a preponderance of positive or negative responses, or a clear mixture across the users). Such trends can then be used to help guide the focus group questioning which should occur after the written surveys but as soon as possible to ensure soldiers' impressions are still fresh in their minds.

Specific Concerns

I acknowledge your concern about introducing unplanned questions into after-action meetings and focus groups. Nonetheless, while the potential exists for adding, on the fly, questions that do not adhere to the best practice guidance, we should not limit ourselves to a strict set of questions articulated in the test plan, in light of the above-described guidance on the conduct of focus groups. Since focus groups are to be diagnostic in nature, and questions formed on the basis of the responses from the written surveys, it would be too constraining to limit the interviewer to only those questions established in the test plan – we want the focus groups to provide us information through a dialog with the soldiers; a stilted overly-structured line of questioning will limit the effectiveness of the focus groups. Therefore, you should make certain the interviewers are properly trained so that they will adherence to best practices (e.g., neutrality in questioning, etc.), while maintaining the flexibility to ask questions 'on the fly' as necessary to obtain more insight into why the soldiers answered the written surveys in a particular way.

You also expressed a specific concern with questions asked to participants to render subjective opinions in areas where they may have no experience. While I agree with you that we should avoid interviewing and surveying soldiers that have no experience on the system under

---

[3] For more information, see R. Barcikowski,"Statistical Power with Group Mean as the Unit of Analysis", Journal of Educational Statistics, Vol. 6, No. 3 (1981).

[4] In fact, this was observed in one recent case where the group commander openly expressed an opinion prior to the group interview which apparently persuaded many of the subordinate soldiers to reply in kind.

test, we do want to survey and interview the soldiers that are newly-trained operators. In fact, the purpose of the survey is to collect their *subjective* opinions in a quantitative and/or diagnostic manner. The fact that they are inexperienced can be accounted for in the demographics data that accompany their response. Assuming we interview and/or survey users of a variety of experience levels, we should analyze the effect of user experience level on the survey responses. This is a goal of the surveys in many cases; therefore, we should not avoid surveying the least experienced operators as part of operational testing. They are the users; so, by definition their responses are valid for the assessment.

One way we can avoid encountering cases that led to your concern is to include in the test plan a survey and interview *administration plan* that articulates who is going to take each survey, who will participate in interviews and focus groups, when will these surveys be administered, and so on. Doing so will avoid cases where soldiers who should not participate in the surveys are included, and will enable a more structured data collection effort to assure the surveys are sampling across the user population.

System Usability/Utility

Given the recent debate over the issue, let me clarify the use of the specific question "would you take this system to war?" Determining whether the users prefer the new system over the old, whether they feel comfortable/safe using the system, and whether the users feel the system will enable them to succeed in combat, are crucial questions that, in my view, should be asked of the users. We should, therefore, not avoid asking questions about system usability and operational utility in a combat environment. Nor should we eliminate questions from test plans simply because some observers anticipate a negative response.

Below are some examples of this type of question that adhere to best practices and should appear in the written surveys instead of asked as a polled yes/no question during focus groups. As always, the choice to use these questions should be commensurate with the goals of the test and survey plan; therefore, not all test plans must use these examples, but I expect many will since we are often interested in the operators' opinions of system utility and usability. Focus groups or individual interviews can then be used to follow up on the particular responses, and elicit a discussion about why the users felt a particular way about the system.

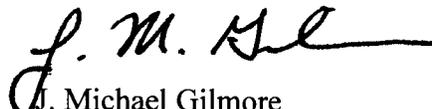Sample written survey questions (not exhaustive):

- I would like to use this system to accomplish the mission. [Likert-scale response – 6 levels ranging from strongly disagree to strongly agree]

- I would take this system to war. [Likert-scale response – 6 levels ranging from strongly disagree to strongly agree]

- The system is easy to operate. [Likert-scale response – 6 levels ranging from strongly disagree to strongly agree]

- Are there any improvements that you would make to the system? [open response]

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

<u>Sample follow-up interview or focus group questions (not exhaustive):</u>

- Some of you answered "strongly disagree" to the statement "I would take this system to war" – would a few of you please provide some reasons for why you felt this way.

- During what portions of your mission or during what activities did the system work particularly well/poor?

- If you could change one thing about the system what would it be?

- Half of you said the system was relatively easy to operate, half of you said the opposite. Could a few of you explain why you answered positively? For those of you that answered negatively, can you explain what task you found to be difficult to accomplish with the system?

I look forward to continuing to work with you to implement my guidance on the use and design of surveys.

J. Michael Gilmore
Director