



OPERATIONAL TEST
AND EVALUATION

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

APR 02 2015

fin
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



scales¹. Additionally, from a statistical power perspective (discrimination ability), the power decreases rapidly when you drop below 4 options. Consequently, response scales should have between 4 and 7 levels and should be validated scales as established in the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) Questionnaire Construction Manual Annex.


Finally, you mentioned using the Hodges-Lehman estimator in addition to the chi-squared test and the sign-rank test I had previously suggested as appropriate. This is of course an acceptable robust estimator derived from the sign-rank test and may be useful when there is missing data. The analysis methods I suggested in my original guidance were not exhaustive and were only intended to provide an initial set of analysis tools. However, I should note that while these tools are acceptable they may not be optimum. If constructed properly, Likert-like scales can typically be treated as continuous data allowing the use of inferential statistical analysis, such as analysis of variance or statistical regression methods. Overall, if we can use parametric methods they are more powerful than non-parametric methods.

A recap of best practices related to your email for closed response surveys includes:

1. Use neutral responses appropriately:
 - a. Neutral responses should be avoided when it is clear that the operator will have some opinion. For example, operators should have an opinion when asked whether “This system is easy to operate.”
 - b. Neutral responses can be interesting and should be included in cases where a direct comparison is being made. For example, operators may have no preference when asked whether “The new system was easier to use than the legacy system.”
2. We should avoid asking questions of operators that do not apply in order to keep motivation high for completing the surveys and to minimize the number of questions each operator must answer.
3. Avoid having a “don’t care” or neutral response outside of the scale. To the extent possible we should avoid “not applicable” response options altogether.
4. Use response scales with between 4 and 7 levels.
5. Use Likert scales or other validated scales as established in the ARI Questionnaire Construction Manual Annex.

¹ For more information see: Preston, C. and Colman, A., “Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences”, *Acta Psychologica*, 1-15, (2000)
Parker, R. et. al., “Reliability of mulit-category rating scales”, *Journal of School Psychology*, (2013)
Lozano, L. et. al., “Effect of the Number of Response Categories on the Reliability and Validity of Rating Scales”, *Methodology* (2008).

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 check assumptions with the data so we cannot guarantee before testing that this is appropriate, but it is an option, which is why Likert scale data is used in the first place.


J. Michael Gilmore
Director