

## Script for the Slides on Comparative Analysis of Ratings Using the Old and New COER

1. Thank you. Good morning Sir, members of the Professional Advisory Council, Chief Professional Officers, PAC Chairs, and guests. The Commissioned Officers' Effectiveness Report (COER) Comparative Analysis includes the methods used, results of and conclusions from the study.
2. First, an appreciation is extended to the officers and raters who participated in the comparative analysis and a number of colleagues for their professional and technical expertise. Most are officers but to obtain objectivity, two are not, one is a civil service employee and the other is in academia.
3. The purpose of the study was to compare the scores from the new revised scale with the old scale, determine the usability of the new, revised version, and identify appropriate intervention in terms of training and education.
4. The old COER Scale had on a single page 18 questions with options to select from A through E or F. The guidance was provided in Manual Circular PHS No. 373.
5. While the new, revised version has eight elements, with 1 through 7 scales and behavioral guidance to help the rater to provide a more accurately reflection of the officer's performance.
6. For the target population, a 2% random sample of officers was drawn. Each was emailed a description of the study and a consent form for their participation and ensure confidentiality was maintained. Most of the forms were received in January. Each of the officers' 2009 raters was emailed a packet with information about the study, the decision memo, the new COER form, and the officer's attachment 1.
7. The data generated from the COER is categorical and hypergeometrically distributed. To examine the significance of association among the dichotomized variables, the Fishers and McNemar tests for the 2x2 contingency tables were used. For the rating point by point pair comparison, the Wilcoxon test for Non parametric independent data was used.
8. The target population consisted of 121 officers. We hoped for a 50% response rate but we got 45%. Of the 54 who provided consents, 50 were scored by the raters and 47 were compared. Three were not included in the comparison because the 2009 COERs were submitted as narratives or were incomplete. We had 36% nonrespondents and 19% nonparticipants.
9. The demographic variables analyzed were sex, rank, racial status, professional category, years of service, organization, isolated hardship assignment, and officer or non-officer rater.
10. For statistical analysis, the variables were recoded into a dichotomized variables, such as, clinical or nonclinical, senior or junior officer, minority or not. The clinical professional category included the officers in the Dental, Medical, Nurses and Pharmacy professional categories. All the variables were dichotomized including rank, years of service, and rater. The results from the demographic analysis are presented in number, proportion and graphs here and in the following slides..
11. The categories, or variables, are in the first column, the number out of 50, and the percent. Of the 50 officers that were scored, 72% of a clinical professional category, 62% had more than 10 years of service, 56% were senior officers, 54% were scored by an officer, 40% were men, 36% were up for promotion and 34% were minority. Nine percent of the officers in the sample were in isolated hardship assignments.
12. A large proportion of the officers received E scores in 2009. The range was from 48% to 86%, with 12 around 80%. Questions 7, 12, 13, 15 and 17 are those that have the F option. Questions 7, 15 and 17, in yellow, are questions regarding supervisory or managerial responsibility

questions. Question 12 is on response, and Question 13 is on growth in skills. I'll mention here that Questions were with the F option were excluded from the pairing test because they are unpaired with the elements in the new COER.

13. The proportion of officers that received a score of 7 using the new COER was lower, with a range of 26% to 36%.
14. The considerations for rating the comparison are listed on this slide. As I mentioned, the questions with the F option were excluded. A to E were recoded from 1 to 5. The new elements were paired with the closest old question. There were no specific match for leadership and interpersonal. The expectation is that leadership should be demonstrated by each officer and the interpersonal were asked in several questions. The overall scores from Element 8 and Question 18 were not included in the averages.
15. In further describing the pairing for comparison, of the 50 officers' scores for the study, 47 old Question scores were paired with the best matching new Element. Three 2009 COER scores had incomplete data. Excluding the questions with F, 31 officers got Es for each of the remaining 13 questions for the 2009 scores. However, seven officers received a score of '7' for each of the eight elements using the new COER scales.
16. The overall scores, Element 8 and Question 18 plotted here for each officer. The officers, 1 to 47, are on the X axis and their score on the Y axis, with Element 8 score, in blue diamonds and Question 18, in pink squares. We see that 41 officers had E on the old and 13 had 7 on the new.
17. This graphs show the officers average old and new scores. That is the averages of the 7 Elements and 12 Questions. Using the new COER, seven officers got perfect scores and 31 using the old COER. As the average old scores dropped, we see more scatter on the lower end.
18. In these scatter plots, the old and new overall scores were plotted against each other with the old scores on the X axis, and new scores on Y axis. The overall scores, Element 8 and Question 18, are the left. The averages are the right. None of the officers were scored higher using the new COER. Since the old question had only two data points, D and E, the scores were averaged.
19. The dichotomized variables using contingency tables were analyzed for association using the McNemar and Fishers Exact test. These tests are ideal for small data sets. The only two that were significant were that women were more likely to be in a clinical professional category and that junior officers were more likely to be have an officer as a rater.
20. The p-values results from the paired point to point comparative test are presented in Table 6. None were significant and ranged from 0.3 to 0.9. I ran a pair of the averages of got 0.5.
21. The tests used are robust measures for small numbers and in the ranks the score for the point to point comparison so assumptions about the distribution is needed. A p-value of a 0.5 is a coin toss.
22. The interpretation of the rating results using the new COER was more random with higher variability. The old and new do not measure the same. The paired questions/elements may be similar but the ratings are different individually. And, the rating using the new is different from the old one.
23. The limitation of the study was the pre-selection, small sample size attributed to the nonparticipation due to being called to active duty within the last year, having no 2009 COER, or had a narrative 2009 COER. For non-respondents, we cannot predict why officers choose not to participate so we do not how their raters would score. Perhaps their scoring would be more randomly distribution, or the officers did not agree with the rater. Some raters, provided comments but it was not requests and not all raters provided it so it was not part of the analysis

but in general most provide positive feedback. Cognitive analysis was not part of this analysis that would consider word count and behavioral questions.

24. The recommendation is to develop an education plan and provide training. To begin this process, an Educational Advisory Panel (EAP) will convene to develop a training curriculum and educational plan. The EAP will consider leadership involvement, accountability of all users, behavioral and cultures changes from inflation, and score based on performance, competency and need for improvement. The final report and the curriculum will be disseminated through publications and presentations. Any recommendations you provide will be appreciated and considered.
25. Thank you for your audience. If you have any questions, you can contact me, or CAPT Plasencia.