CASE STUDY
OF A PROTOTYPE SET OF
BEHAVIORALLY ANCHORED RATING SCALES (BARS)
FOR C2 ASSESSMENT

Jim Murphy
Dynamics Research Corporation
1755 Jefferson Davis Hwy, Suite 200
Arlington, VA 22202
703-412-2812 ext 6747
jmurphy@drc.com

Jock O. Grynovicki, PhD
U.S. Army Research Laboratory
ARL-HRED, Building 459
Aberdeen Proving Ground, MD 21005
410-278-5956
jgrynovi@arl.army.mil

Kragg P. Kysor
U.S. Army Research Laboratory
ARL-HRED, Building 459
Aberdeen Proving Ground, MD 21005
410-278-8824
kysor@arl.army.mil

This page intentionally blank

USE OF THE BATTLE COMMAND PROFICIENCIES CONCEPT IN THE UNIT OF ACTION BATTLE COMMAND EXPERIMENT

ABSTRACT

1. The Command and Control Research Program's (CCRP) recent publication of the "Code of Best Practice for Experimentation" highlights once again the potential of behaviorally anchored rating scales (BARS) as an assessment tool for command-staff team performance in the execution of battle command. Interest in BARS as a performance appraisal technique has grown steadily since first proposed in 1963 (Smith and Kendall). The BARS methodology was originally proposed as an appraisal instrument for individuals but has since been applied to team environments. In 1991, the U.S. Army funded research for the development of BARS to support aircrew training (Grubb *et al*.). In 1999, a comprehensive framework was developed for assessing C4ISR performance in which commander and staff proficiencies would be observed and assessed using carefully developed BARS (Leedom). The intention was to link the assessments to other force effectiveness measures to determine the overall effectiveness of the human-technical system. That proposal led to an ARL funded effort and the development of a *prototype* set of BARS designed to assess the team proficiencies of a commander and staff in the execution of "battle command." This paper describes the use of the prototype BARS in a recent battle lab experiment, and sheds light on future efforts to develop BARS with in the context of Joint services experimentation.

BACKGROUND

2. In 1999, the Army Research Laboratory (ARL) began to develop a framework for accessing C4ISR performance (Leedom). At that time, the Army's Force XXI digitization programs were in the early stages of a development process that would eventually lead to the distributed C4ISR concepts underlying the Army's Future Combat Systems (FCS) development process. ARL noted that although technology supported the C4ISR systems, the human decision–making process was governed by the socio-cognitive limitations, strategies, and interactions of commanders and staff officers. At the center of the C4ISR system would lie a distributed human decision-making process—a process that could be supported by technology—but a process that would still be governed by the socio-cognitive limitations, strategies, and interactions of commanders and staff officers. At the time all of DoD was increasing research and development efforts in the C4ISR disciplines. ARL described three significant challenges: (1) determining the appropriate mix of investments in the human and materiel components of battle command, (2) determining how best to integrate the components, and (3) determining the relative returns on investment (ROI) associated with each human and materiel R&D program. A common assessment framework was proposed to address the three challenges—one that would allow each type of investment to be assessed against its contributions to decision-making performance in battle command (Leedom).

3. At the heart of the research concept was an assessment methodology to identify key battle command team or organizational proficiencies. The development process was to ensure that (1) the definitions and sub-categories of the proficiencies would be operationally defined with an acceptable degree of agreement and precision, (2) the descriptions of the proficiencies would facilitate focused observation and assessment

during battle command operations, (3) the products would be able to be empirically correlated with other aspects of C4ISR system performance (e.g., MDMP task completion, battle staff products, battle outcome), and (4) the assessments would provide a reasonable level of diagnostic feedback regarding specified combat development shaping factors. The approach was to identify key battle command team proficiencies, then develop sets of BARS for each proficiency. Given the complexity of the cognitive and behavioral dimensions of battle command, an incremental, six-phase BARS development process was designed to achieve the highest possible degree of validity and reliability for the resulting behavior scales (Leedom). Figure 1 outlines the six-phase development process.

STEP	OBJECTIVE	APPROACH
1 Theoretic Definition	Expand proficiency definitions into a detailed discussion of proficiency goals and pathways by which each shaping factor influences behaviors (construct validity)	In-depth review of existing research literature
2 Behavioral Description	Identify and rank-order sets of behavioral markers that characterize unacceptable, minimally acceptable, and superior levels of proficiency along each dimension (face validity	Critical incident analysis, based upon data available from previous AWE events, NTC/JRTC rotations, and CALL database
3 Observer/Rater Facilitation	Develop observer/rater training materials that focus attention on key aspects of each proficiency and facilitate systematic assessment of battle command behaviors during an exercise (rating validity)	Lecture material, supplemented with case studies and sample observation exercises fro developing inter-rater reliability
4 Field Demonstration	Conduct actual applications of the C4ISR framework (and associated BOS/BARS scales) within on-going battle staff exercises and field experimentation: (1) organization of experimentation issues, (2) collection of performance observations, and (3) analysis of findings and insights	Participation in strike force experimentation events (TRAC-led), Command Post of the Future experimentation events (DARPA-led), and Command Post XXI experimentation events (CECOM-led)
5 Database Development	Develop a central repository for C4ISR combat development assessment, organized by battle command proficiencies cross-walked against both shaping factors and higher levels of C4ISR system performance (criterion validity)	Collaborate with both CALL and TRADOC Battle Labs to organize a central repository of battle command proficiency assessments
6 Product Handoff	Document battle command proficiency findings and insights in appropriate forms for handoff to training developers, materiel developers, and combat development centers	Technical reports (published either by ARL or in collaboration with TRAC, TRADOC battle Labs, DARPA, and CECOM

Figure 1. Research Steps for Developing a C4ISR Assessment Framework

4. The basic concept document, which was based on extensive field observations and a comprehensive literature review, identified 17 commander-staff team proficiencies across the range of battle command activities. For ease of use, the 17 proficiencies were clustered under four broad organizational performance goals. This initial effort set the conditions to begin the six-phase development process outlined in figure 1.

5. An ARL-directed project team accomplished the detailed work on steps 1, 2, and 3. Initially, the team developed a detailed doctrine, tactics, and procedural description for each proficiency. The team then decomposed each into its component sub-proficiencies, yielding 62 sub-proficiencies. Each sub-proficiency was then described in terms of the behaviors an observer might would see when the team was engaged in this proficiency. The written behaviors attempted to capture three levels of commander-staff team effectiveness. At one end of the scale, behaviors were drafted to describe actions or reactions of the commander-staff team that, on the basis of the project team's collective

experience, were expected to achieve the highest effectiveness for the organization. These were generally the easiest to recall or envision. The second set of behaviors attempted to describe actions and reactions that were effective in accomplishing the task, but only minimally so. These were the most difficult behaviors to describe. The third set were the behaviors that were clearly ineffective in achieving an effective result for the organization. These were relatively easy to describe, as well. The project team's report is in a paper titled "Development of a Battle Command Staff Proficiency Measurement System I."

6. The product of Steps 2 and 3, referred to as the Battle Command Proficiencies BARS, has been used by ARL to develop data collection plans for TRADOC Analysis Center (TRAC)-sponsored battle command analysis efforts. These include: (1) the Division Capstone Exercise I, (2) Division Capstone Exercise II, and (3) the Army Transformation Experiment 02. Recently, the battle command proficiency BARS were used in an operational-like setting, an experiment at the Battle Command Battle Lab – Leavenworth (BCBL-L). The resources available and the time allocated to the ARL team during the experiment permitted only a relatively limited "proof of principal."[1] The team's objective was to demonstrate that the prototype BARS could be used effectively in an operational setting by persons with military domain expertise. It was not possible to assess, as indicated in Steps 1 and 2, above, the construct and face validity of the BARS.

UNIT OF ACTION (UA) BATTLE COMMAND EXPERIMENT

7. The experiment, conducted 30 January – 7 February 2003 at Fort Leavenworth, KS., was designed to provide insights into five major questions related to the Army's Objective Force. The questions were:

1 – Does the UA Brigade command and staff structure enable battle command?
2 – Does the Recognitional Planning Model (RPM) support planning, execution and decision-making across the full spectrum of military operations?[2]
3 – Is the UA Brigade Staff able to conduct distributed planning?
4 – How does Commander's Intent facilitate decentralized execution?
5 – What are sufficient characteristics of collaboration tools for the commander?

ARL TEAM PARTICIPATION

8. The U.S. Army Research Laboratory team's participation focused on providing input to question 1 and on determining the usability of the prototype BARS by persons with military domain expertise. The prototype BARS were suitable for responding to the BCBL-L's first research question only.

SCOPE OF THE EXPERIMENT

9. As stated in the Experiment Directive, the scope of the UA BC experiment was to focus on the new UA Brigade level staff structure as the commander-staff teams

[1] The BCBL-L integrated five separate research teams from the Collaborative Technology Alliance (CTA) [www.ctac2003.com] into the experiment design. The number of teams required an efficient allocation of time among the participants, and this limited the number of research objectives each team could achieve.
[2] The RPM was presented by Dr. Gary Klein to the 1999 CCRTS at the Naval War College, Newport, RI, 29 June – 1 July 1999.

conducted multiple operations while using a new decision-making process (DMP).[3] The new DMP is a commander-centric, execution-based process designed to support UA-level commanders operating in time-constrained environments where using the full MDMP may not be practical. The directive stated that the insights gained would help refine the core functions of Battle Command performed by the commander and provide preliminary data on the UA structure and the new DMP.

FUTURE STAFF AND FUTURE PROCESSES

10. The details of the exercise design and the actual participant manning has been provided in the BCBL-L's final report. Figure 2 is a composite graphic intended to provide a framework for the discussion that follows. In the upper left corner, the graphic shows the conceptual staff functional responsibilities. A small image of the RPM is in the upper right corner. In the lower left is a diagram of laboratory spaces with the commander, deputy commander, and functional staff members arrayed at tables. Along the bottom are the principal doctrinal and procedural references for the UA brigade staff functions. On the left is the list of UA brigade command and staff positions for the experiment.

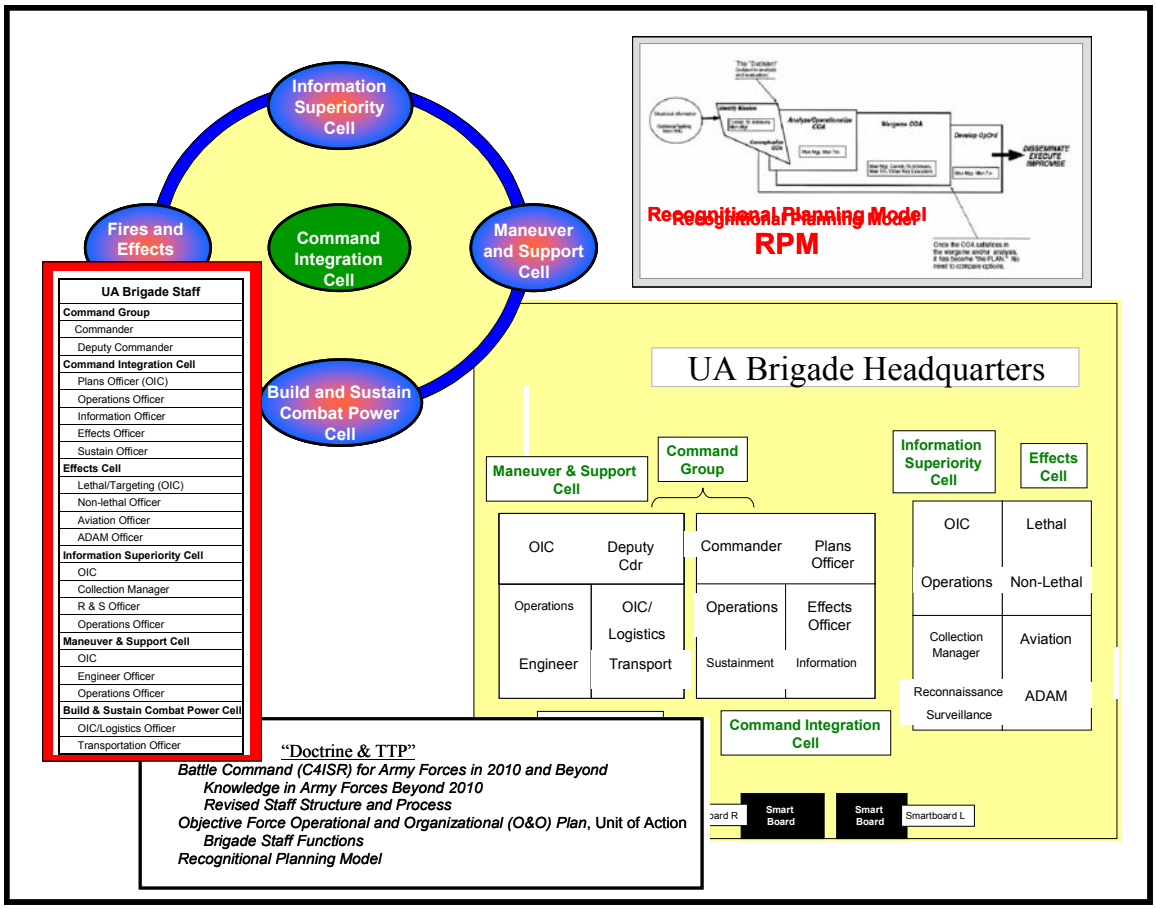

Figure 2. Future Staff, RPM Diagram, UA Staff Positions, and Doctrinal Framework

ROLE PLAYER PARTICIPANTS

[3] Subsequent to the publication of the Experiment Directive, the term "new decision-making process (DMP)" was changed to the "Recognitional Planning Model (RPM)."

11. The UA brigade command group and staff was comprised of active duty officers and highly qualified recently retired officers. The level of experience was important for four reasons bearing on battle command proficiency. First, the future staff structure is so new that detailed descriptions of the staff members' duties had not been written at the time of the experiment. Second, the capabilities of the Objective Force equipment and weapons systems are new and the participants had not had time to absorb the performance characteristics and the tactical effectiveness of the systems and ammunition. Third, the organization of the Objective Force units, employing entirely new equipment was new to most of the active duty participants. Fourth, it was necessary for the team to learn the details of the RPM process and then to practice so that the steps became as familiar as the Military Decision Making Process (MDMP). The core group of the retired officers had participated in similar Objective Force activities several months before the experiment. Their experiences with the Objective Force doctrine, organization, and materiel enabled them to tutor the other participants as necessary, but as with the rest of the staff, this was their first exposure to the RPM.

Concepts and Procedures Bearing on the ARL Analysis Plan and Data Collection Plan

12. The following documents were provided to the participants to develop a conceptual framework related directly to the sequence of command and staff action in the Objective Force environment:[4]

- TRADOC Document (No. Unk.): *Battle Command (C4ISR) for Army Forces in 2010 and Beyond*, Version 4, 14 June 2002
- Annex D, *Knowledge in Army Forces Beyond 2010*, to TRADOC Document (No. Unk.)
- Annex B, *Revised Staff Structure and Process*, to TRADOC Document (No., Unk.)
- TRADOC Pam 525-3-90 *Objective Force Operational and Organizational (O&O) Plan*, Unit of Action, Change 1, 25 Nov 2002.
- Annex D, *Brigade Staff Functions* to TRADOC Pam 525-3-90, Ch 1
- With respect to the *Recognitional Planning Model*, a detailed, 22 page description of the RPM and the sequence of actions to execute the process. v2.2, 12 Jan 03.

Conditions Bearing on the ARL Analysis Plan and Data Collection Plan

13. The following environmental conditions bound the ARL analysis:

- Staff was co-located in one space, not distributed in six vehicles as envisioned in the Objective Force design.
- Simulated C4ISR system was able to provide only a fraction of the functionality that is planned for the Objective Force, to include the collaboration tools necessary to operate in six separate vehicles.
- NCO and operator–level structure necessary to facilitate staff interaction and products was not included in the player structure.

[4] These documents and other relevant readings for the Objective Force, as well as current research reports and articles on naturalistic decision-making were provided to the participants in paper form before the experiment. The documents were disseminated to the entire experiment staff on CD-ROM at the conclusion of the experiment. See BCBL-L CD-ROM *UA Battle Command Experiment*, Fort Leavenworth, KS, 27 Jan-7 Feb 03.

- An intelligence collection, analysis, and fusion capability—comprised of a simulation, personnel, and intelligence systems—was not available to support the staff during the experiment, thus limiting realism in the scenario.
- Of necessity, staff procedures were (1) guided by Objective Force concepts, (2) shaped by the RPM procedures, (3) elaborated on by the UA brigade commander, and (4) executed by individuals relying on their current era training and experience.

The Analysis and Data Collection Challenge

14. Given the understandable limitations in creating an experiment environment that accurately simulated an envisioned future environment, the ARL analysis is valid for conditions as they existed during the experiment only. We believe, however, that the analysis related to the command and staff structure supports the insight that the structure will support battle command in the Objective Force environment.

15. The challenge for the ARL team was to develop an analysis plan and a data collection plan which were flexible enough to adapt to the projected future conditions, but specific enough to generate data relevant to the experiment objectives, particularly under the actual experiment conditions. In current force organizations, "Human Behaviors of Battle Command" unfold in an environment rich with doctrinal and procedural detail. The Battle Command Proficiencies (BCP) concept is rooted in principles of military leadership and organizational behavior and does not rely on explicit written procedures. The BCP sub-proficiencies are concept-based and are not written in terms of specific technology. The BCP concept is expected to be applicable to commander-staff team performance for years into the future. Parenthetically, it is clear that a number of proficiencies are closely tied to digital systems and the related human-computer interface (HCI). Thus, to remain relevant, the basic BCP BARS need to be supplemented by technology-specific BARS, and that these BARS need to be up-dated as new interfaces are introduced.

16. To be described in detail later, the data collection plan centered on each experiment participant using the BCP BARS to assess the degree of battle command proficiency attained by the commander-staff team by the end of the experiment.

DESCRIPTION OF THE BATTLE COMMAND PROFICIENCIES

16. While complementary to the individual and collective tasks in TRADOC Mission Training Plan (MTP) training materials, the BCP combines these and other tasks identified in research literature into sets of individual and team cognitive activities

necessary to integrate battle command functions. The 17 proficiencies are further decomposed into 62 more finite activities, which are referred to as sub-proficiencies. The proficiencies are defined in terms of a doctrine-based description, performance standards, and general procedures and techniques. Figure 3 is a graphic outline of the goals, proficiencies, and sub-proficiencies.

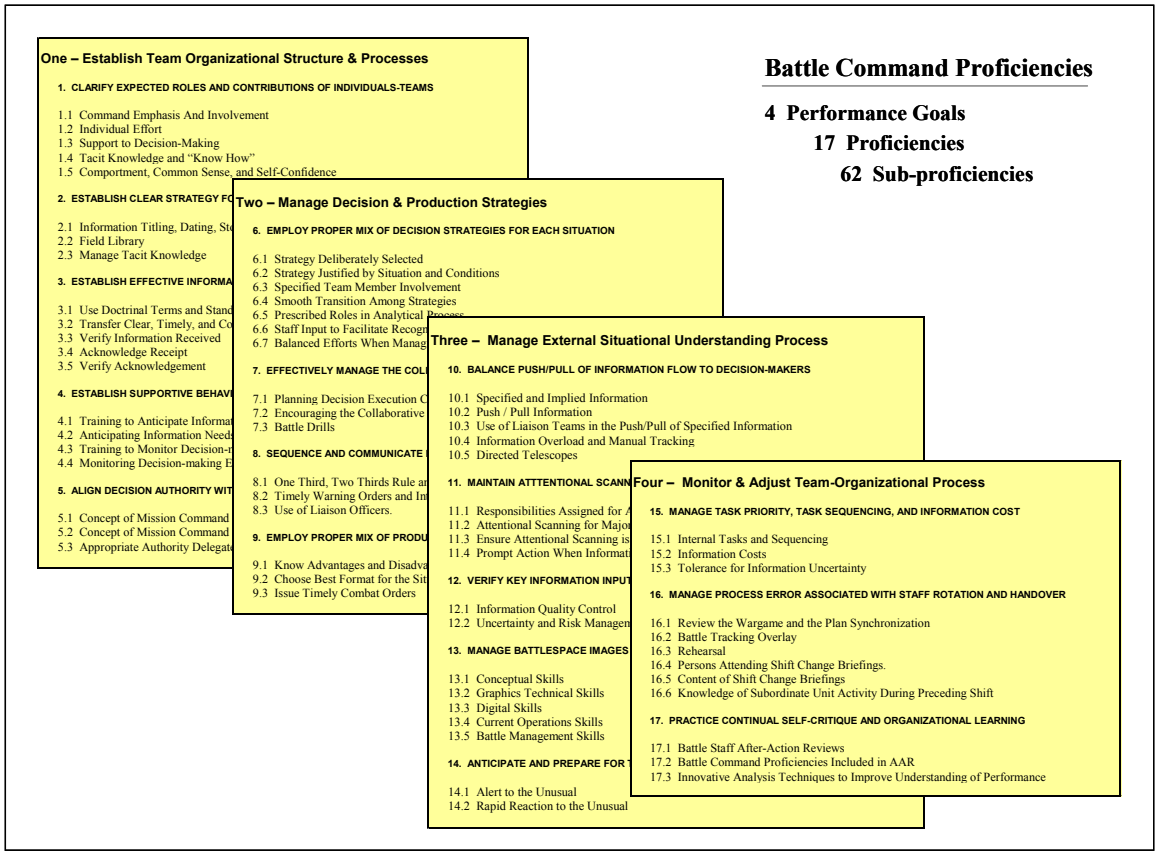

Figure 3. Outline of Prototype Battle Command Proficiencies

18. Most significant in terms of applying and using the BCP, three behavioral descriptions are provided for each of the 62 sub-proficiencies. The content of the descriptions ranges from the most effective behaviors in accomplishing the tasks to the least effective behaviors. Thus, a person observing a commander and staff in an operational setting is able to note the behaviors among the group, find the behavioral description on the scale closest to the behaviors observed, and assess the relative effectiveness of the group in the performance of that task with respect to the overall objective of achieving effectiveness in battle command.

A PROTOTYPE

19. The current set of BARS, configured for the experiment as a survey, is a prototype assessment tool. The prototype was developed in 2001 as Steps 2 and 3 of a projected six-phase development project. The UA BC experiment is the first opportunity ARL has had to examine the BCP, specifically, the BARS in an operational setting.[5] The formal

[5] A key word search on "Google.com" using "behaviorally anchored rating scales" and "development" generates many resource descriptors on the subject. Civilian organizations typically use BARS to differentiate between unacceptable, acceptable, and outstanding behavior or between entry/novice,

BARS development process is highly rigorous, as well as manpower and time intensive. Absolutely key is the intensive involvement by experienced practitioners of the work being studied. BARS describing battle command must be developed by military officers and senior NCOs. The subject matter experts are guided through the development process by analysts who ensure necessary psychometric rigor. The subject matter personnel need to be directly involved for several weeks.[6]

20. As envisioned in the original BCP BARS development plan, the involvement of the officers and senior NCOs required to execute the formal development process does not occur until Step 4. For this reason, the development analysts, comprised of retired military officers experienced with the formal BARS process, approximated several of the key development methods designed to achieve validity and interrater reliability in the outcome. Important to note, however, the four performance goals and the 17 actual proficiencies had been identified as the precursor to Step 1 and after detailed study and analysis of five knowledge domains.[7] Thus the starting framework provided to the development analysts is considered to have been very carefully developed.

RECOGNIZABLE AND RELEVANT

21. Among the features of properly constructed BARS is that the description of the task(s) and rating scales for each task should be immediately recognizable and understood by persons working in the domain for which the tasks have been prepared. We would expect a group of Army officers reviewing a set of BCP BARS to quickly recognize the relevancy of the content, and the effectiveness levels across the three behaviorally anchored scales. Figure 4 on the following page illustrates this point. The upper quarter of the figure describes the sub-proficiency and states in question form the key elements required to demonstrate it.

intermediate, and expert capability levels. Once a set of BARS has been established by for a work unit, it can be used to provide peer and self-evaluation, set development plans, guide selection processes, and aid in the orientation of new members.

[6] Included in the report to the BCBL-L was a concise description of the formal BARS development process.

[7] Leedom (1999). The literature study encompassed (1) Team Performance and Team Training, (2) Military History and Classic Case Studies in Battle Command, (3) Cognition and Naturalistic Decision Making, (4) Organizational Psychology Emphasizing Adaptive and Learning Organizations, and (5) Complexity Theory and Sociological Application to Military Operations.

8.1 <u>One Third, Two Thirds Rule & Planning Timelines</u>. Does the battle staff complete planning and issue the operations order within the one-third, two-thirds guideline? Does the battle staff develop an internal planning timeline very soon after receipt of mission and adhere to it? Does the battle staff subsequently coordinate timelines with its higher headquarters, and issue an expanded planning-briefing-rehearsal timeline to its subordinate units?	
7	The commander and staff are well-disciplined to execute their planning requirements within the 1/3 – 2/3's allocation of time. The staff first determines the amount of time in the 1/3 allocation, then determines 1/4 of the time, and allocates the 1/4 to the MDMP planning tasks. Once planning is underway, the XO or S3 coordinates with HHQ to determine the times for the brief back to HHQ and the HHQ's rehearsal. The S3, with the CO's approval, issues the unit's own briefing and rehearsal times to the subordinate units.
4	The commander and staff normally executes their planning requirements within the 1/3 – 2/3's allocation of time. The staff first determines the amount of time in the 1/3 allocation, and further allocates it to the planning tasks. The staff tends to wait for the higher headquarters to announce its briefing and rehearsal schedule. The S3, with the CO's approval, issues the unit's own briefing and rehearsal times to the subordinate units.
1	The battle staff invariably overruns the allocated time, taking up to 1/2 the time available. The staff simply has difficulty completing all the steps within the time intervals they initially determined. The unit tends to wait for the higher headquarters to announce the briefing and rehearsal schedule, and as a result, frequently does not begin to coordinate this schedule until they have completed their operations order. This causes other units to have to delay or reschedule

Figure 4. BARS Content Must Be Recognizable as Relevant to the Tasks

The paragraphs at 7, 4, and 1 are behavioral descriptions of the sub-proficiency. The behavior at 7 was assessed by the project analysts as the most effective in accomplishing the task. The behavior at 4 is considered minimally effective, and the behavior at 1 is considered to be least effective. In fact, it is ineffective. Behavioral anchors were not proposed for the levels of effectiveness that would relate to positions 6, 5, 3, and 2 in the continuum. The persons using the scales are instructed to use these numbers to indicate behaviors that lie between 7 and 4, and 4 and 1.

Some Academic Language

22. Not all proficiencies were stated in such clear operational language. A number of the proficiencies and sub-proficiencies were expressed in more academic terms. Academic phraseology is acceptable in a prototype, albeit somewhat problematical when working directly with soldiers. Future versions of the BCP need to be expressed entirely in operational terms with which the soldiers are completely familiar.

USES: TRAINING, UNIT SELF IMPROVEMENT, READINESS,

23. Although developed specifically for purposes of C4ISR R&D assessment, BCP BARS have the potential to support three other areas (training, unit self improvement , readiness)

Training Inventory

24. BCP BARS could be used by trainers to assess a unit's current level of battle command proficiency and develop an action plan for the unit to improve. The trainer would observe the staff, identify behaviors, and mark his rating sheet. The behavioral anchors encompass individual skills, effort, experience, teamwork, and leadership. The trainer would have to parse out which components were present in a team's behavior to determine what training—skill training, leadership training, or both—would be necessary to improve team effectiveness. The BARS would facilitate establishing the proficiency level the team had attained, and the increments of effort necessary for improvement.

Unit-initiated Self-critique and Organizational Learning

25. The unit could use the BARS for unit-initiated self-improvement. This is an ideal use of the BARS. It would spring from the unit's own motivation and rely on the members to be objective in assessing their own proficiency. It would be highly complementary to the well-established practice of After Action Reviews (AARs). In effect, this was the technique used to collect data on the BCP BARS during the experiment.

Readiness Assessment

26. Formally developed BARS could constitute a supplement to current processes for assessing unit readiness. BARS focused on battle staff proficiencies could provide a framework for readiness assessment and written standards with which to estimate unit preparedness to exercise complex battle staff processes.

VALIDITY AND INTERRATER RELIABILITY

27. The abbreviated development process used for the BCP BARS limited the degree to which the behavioral anchors scales could meet accepted standards for psychometric validity and interrater reliability. Given this, the ARL team was prepared for noticeable variability in the ratings of each sub-proficiency by the officers participating in experiment. The analysts simply did not know the degree of rating variability the responses would reflect. We define three types of rating variability below. We use these measures of rating variability later in presenting the results.

VARIABILITY – THREE TYPES

Individual Rater Variability

28. An evaluator whose ratings across all proficiencies were within two or three scale numbers (e.g., all assessments were a 4, 5, or 6) would have a narrower individual rating variability than a person who rated proficiencies at five different levels of command staff team effectiveness.

Group Rater Variability within a Proficiency

29. This refers to the distribution of ratings by all evaluators within a single sub-proficiency, for example, how all evaluators assessed sub-proficiency 3.1. When the ratings made by all evaluators fell in a range of three adjacent scales (e.g., 4, 5, and 6) the group rating variability in this sub-proficiency would be relatively narrow. When the group's assessments fell in five or six adjacent scales, the group's rater variability would be much larger. The greater the group rating variability within a proficiency, the more likely the actual sub-proficiency or the behaviors, or both are not valid reflections of actual tasks being observed. A wide distribution of ratings is mitigated somewhat if the preponderance of ratings fall within two or three effectiveness scales and the remainder are distributed in ones and two such that the histogram resembles a steep symmetrical or skewed curve. This type of rating variability relates directly to the psychometric concept of interrater reliability.

30. The total number of ratings by all evaluators in each proficiency can be shown graphically in a histogram. In the ideal case, all raters would reach the same assessment of the staff's proficiency and the histogram would reflect all ratings in one scale. In the more likely case, the ratings were expected to be spread across several scales with a

preponderance of ratings matched to two or three scales. The most populated scales constitute a center of mass.

Group Rater Variability Across the Set of Proficiencies

31. Variability across the set of proficiencies is the degree to which the centers of mass for the entire set are spread across a range of scales. The research analysts had no preconceived notion of what rating variability across the set of proficiencies indicates or what variability should be expected. It was felt, however, that group variability of one scale—meaning all centers of mass fell on the same level of the scale—was improbable. It was also felt that a pattern of rating variability might suggest a particular strength or vulnerability. For instance, Proficiency 6, "Employ Proper Mix of Decision Strategies for Each Situation," has three sub-proficiencies. The inferences to be drawn concerning the commander-staff decision processes are very different depending on whether the pattern of centers of mass for these three sub-proficiencies falls to the left or right of the mean. Falling to the left suggests the commander-staff process is relatively well developed. Falling to the right of the mean suggests that the unit's decision processes are (1) flawed, (2) need careful analysis, and (3) need appropriate corrections. The analysts decided that they would consider the group's rating of a sub-proficiency significant with respect to agreement if 66% of the evaluations fell in two adjacent cells or a single cell. In the summary of results, we include only the sub-proficiencies which reflect the 66% criterion.

PROOF OF PRINCIPLE

32. The salient point is that although the BARS are an excellent initial framework, they are not presented as a statistically reliable and valid instrument for the assessment of command-staff team proficiencies. Thus, the ARL team viewed their use in the Unit of Action Battle Command CEP experiment as a "proof of principle" that persons with military domain expertise in an operational setting could effectively use the BCP BARS.

METHODOLOGY

PREPARATION

33. The ARL team understood that the experiment's full schedule and many research objectives would limit the time the analysts would have directly with the experiment commander and staff participants. The team also recognized that the participants would have limited time to read materials prepared to orient them to the Battle Command Proficiencies project. Notwithstanding, the team extracted relevant sections from the original Battle Command Proficiencies Report to provide the commander-staff raters a BCP orientation booklet.[8] The booklet was packaged in a two-part spiral binder titled:

> "Behaviorally Anchored Rating Scales for the Battle Command Proficiencies:
> Part I – Descriptions, Standards, and TTP (96 pages)
> Part II – Observational Focus & BARS" (54 pages)

The ARL team briefed the commander-staff team prior to the rehearsal phase on the project and the data collection method to be used.

[8] *Development of a Battle Command Staff Proficiency Measurement System I.*

SIMPLE METHOD

34. The BARS technique is generally used in military applications as a training assessment tool by instructors observing training. But as noted earlier, the BARS technique is routinely used in other organizations as a means for team members to assess their own proficiency. For example, the 17th proficiency, "Practice Continual Self-Critique and Organizational Learning," emphasizes the after action review (AAR) process, and recommends the use of the proficiencies as a framework for self-critique and capturing information to be used in unit-initiated AARs. With that in mind, the ARL analysts asked if the commander-staff participants could be integrated into the data collection using the BARS, essentially as a survey. The staff directing the experiment approved the request.[9]

ORIENTATION AND PRE-SURVEY

35. Following the briefing, the ARL team distributed the BCP package to the commander and staff participants, and distributed also the Battle Command Proficiency BARS instrument. The instrument was a sub-set of 35 of the original 62 sub-proficiencies. The 35 selected sub-proficiencies were oriented toward the eight major proficiencies (tasks) which the ARL team was most confident would be observable during the execution phase of the experiment.

The BARS Survey Instrument

35. The prototype BARS are comprised of a considerable amount of text. A challenge for the research analysts was formatting the instrument to retain all necessary text at a reasonable font size while providing sufficient white space for the participant to easily work through the survey. The compromise was to present the survey on 11"x17" sheets. Figure 5, on the following page, shows an example of one of the BARS survey sheets. Each sheet contained the title and a description of the proficiency at the top. Immediately below is a description of the first sub-proficiency, and immediately below that are the seven numerical anchors. Notice that the behavioral anchors are below scale numbers 7, 4, and 1. Figure 5 shows three sub-proficiencies.

Pre-survey

37. The ARL team distributed the BCP BARS survey instruments to the participants at the end of the training phase, and prior to the rehearsal. The pre-survey was intended only to allow the raters to become familiar with the instrument. They were asked to provide their ratings of their staff's proficiency at the point where all training and rehearsal had been completed. This required the raters to read and consider each proficiency, sub-proficiency, and behaviorally-anchored scale prior to the start of the execution phase of the experiment.

38. The commander-staff raters were instructed to read the description of the sub-proficiency, annotated in Figure 5, then read the three behavioral scales. If the behaviors they had observed during the training and rehearsal were similar to one of the behavioral anchors, they were to circle the number above the text. If the behavior was of a more proficient nature than that described in anchor 1 or 4, but not quite as high as the next

[9] In retrospect, the consent by the experiment director to permit the commander-staff participants to complete the BCP BARS survey was analogous to Unit-initiated Self-critique and Organizational Learning technique.

anchor to the left, they were to select one number-only scale between the two behavioral anchors. Which numerical anchor they chose depended on degree. It was also possible for the respondent to circle 7, 4, or 1 and add additional descriptive phrases to describe a behavior not included in the anchor, but believed to be equally effective.

39. The pre-surveys were collected in the 3 days after the execution phase commenced. The research analysts extracted the rating responses from the individual surveys onto aggregate forms to identify trends that might indicate erroneous understanding of the instrument. The analysts did not intend to use the pre-survey data. The analysts reviewed the data primarily to determine variability within an individual rater's responses and group rating variability within a proficiency.

10. BALANCE PUSH/PULL OF INFORMATION TO DECISION-MAKERS

OBSERVATIONAL FOCUS The emphasis in this proficiency that the team has a clear mental picture of what information is relevant to HHQ and subordinate units in a given situation, and that it already has in place the reporting procedures for reporting to HHQ and for receiving reports from subordinate units. This should minimize the necessity to pull information from a subordinate unit, and similarly, the necessity to have to respond to pull requests from HHQ. The unit TACSOP should reflect an integrated approach to information flow. If the unit has not described its information TTP in writing, the observer should see persons taking the initiative to push appropriate information as it is received—such as SPOT reports of significant contact—or as it is scheduled to be sent, such as twice-daily operation summaries and intelligence summaries. The observer should see a minimum of information pull. When the operational pace quickens, and the staff is being overloaded with information, the observer should expect the senior staff members to tighten the filters for the information they want their subordinates to pass to them for consideration.

Liaison teams should be fully aware of their information push and pull responsibilities. Normally, they will be focused on ensuring the battle staff is apprised of all relevant information bearing on their parent unit. If the commander chooses to use a "directed telescope," the tasking should be treated as regular mission, and should be fully coordinated and synchronized. The following outlines specific proficiency elements to be observed:

10.1	Relevant Information. Do the battle staff members understand what information is relevant in the present situation? Has the unit developed procedures to ensure the junior staff members pass only relevant information and other exceptional, time-critical information to the commander or designated senior staff members?						
	7 The command's TACSOP lists the types of information that generally becomes relevant in a tactical situation, and the officers and NCOs on the staff have a clear understanding the types of information that are relevant in a given situation. The operators and NCOs in the staff sections are well-drilled in routing relevant and exceptional information to designated senior staff members, and other important but not immediately relevant information to other designated staff members.	6	5	4 Although the TACSOP does not provide a description of relevant information, the key staff members have a good idea what is relevant in a given situation. Operators and NCOs in the sections understand the concept of relevant information, but have not been given explicit guidance routing it within the staff.	3	2	1 Generally, none of the staff members are clear on what information is relevant in a given situation. Operators and NCOs in the sections tend to pass all information to the senior person in the section
10.2	Push / Pull Information. Does the staff understand what information should be pushed and to which individuals or commands and under what circumstances? Does the staff ensure that this information is pushed on a timely basis? Does the staff understand what information should be pulled and from which individuals or commands and under what circumstances? Is the staff proactive in pushing and pulling the information?						
	7 The document also lists by title the reports that are to be pushed and pulled. Persons responsible for pushing and pulling also identified. All persons working in TAC have been trained in "information and reports management," and are alert to the requirements in the TACSOP. During operations, the battle staff members are alert to ensuring their information is pushed as necessary. In order to confirm that the information being requested from subordinates is sufficiently necessary as to justify the effort and the distraction, staff officers discuss the request with counterparts to get a "peer check."	6	5	4 The command's TACSOP lists by title the reports that are to be pushed and pulled, but persons responsible for pushing and pulling the reports are not identified. The XO and other senior battle staff members have oriented the staff to the reports and to their actual routing during operations. During operations, the staff is conscientious about sending push type information. Though careful not to request too much pull information, the staff has no peer check to act as a governor on excessive use this technique.	3	2	1 The command's TACSOP does not describe reporting in a manner that provides guidance on push and pull information. The XO and other senior battle staff members have oriented the staff to the reports and to their actual routing during operations. During operations, the staff is frequently forgetful and inattentive to timeliness for submission of required (push) reports to HHQ. The staff tends to ask for excessive amounts of pull information from subordinate units, and frequently asks for information during enemy contacts when the subordinate staff needs to be able to concentrate on the situation
10.3	Use of the Liaison Teams in the Push / Pull of Relevant Information. Does the command use the liaison teams located in its TOC to assist in the push / pull of relevant information? Has the command briefed its own liaison teams on their roll in expediting the pus / pull of relevant and exceptional information?						
	7 The TACSOP lists the responsibilities of liaison teams assigned to the TOC. The TACSOP is highly specific in describing in the information the liaison teams are responsible for acquiring and for providing to their organizations. One of the senior members of the TOC staff is assignee to brief liaison officers on their responsibilities. A copy of the TACSOP is provided to the liaison teams.	6	5	4 The TACSOP lists the responsibilities of liaison teams assigned to the TOC. A copy of the TACSOP is provided to the liaison teams. No extra effort is made to ensure the liaison teams are well-oriented to their information role.	3	2	1 The unit TACSOP has no section addressing the reporting responsibilities of liaison teams. The liaison team is not oriented to its information responsibilities upon arrival at the command post.

Cont'd

Figure 5. Example of a BCP BARS Assessment Instrument

40. Only one commander-staff rater had a lack of individual rating variability. This participant assessed virtually every proficiency as a level 7 (the highest proficiency level). The ability to achieve the maximum proficiency level was rather an astonishing accomplishment for a group that (1) had just formed, (2) was exercising staff procedures using a thinly detailed staff organization tailored for a projected year 2015 environment, and (3) was comprised of active and retired officers. This situation might be considered as a classic case of a person who does not want to say anything that might be construed as a criticism of his seniors. One of the team members spoke with the rater concerning his

assessments. The rater responded in a manner to indicate he sincerely believed the assessments he made were the commander-staff team's actual level of proficiency.

41. The ARL team also noted a clear trend toward considerable group rating variability within a proficiency. The variability was considerably greater than the team had expected. The team decided that if the opportunity were to present itself to talk to the commander-staff raters as a group before the final survey, they would be prepared to do so. The team wanted to lay out the spread of assessments in a PowerPoint presentation slide, pointing out the large rating variability, and then encourage the participants to read the behaviorally-anchored rating descriptions carefully before concluding their rating for each sub-proficiency. In fact, this opportunity did present itself at the conclusion of the experimental trials, and before the post survey was taken.

Post Survey

42. The command staff raters completed the post survey the morning after the final experimental trial. At this point, the Unit of Action staff had worked intensively for 4 days. The ARL team had the opportunity to brief the commander-staff raters before they completed the post survey. The analysts stressed the importance of objectivity and reading the behavioral anchors carefully. The analysts illustrated four examples from the pre-survey of group rating variability within a proficiency. Two examples portrayed a very narrow dispersion of ratings, and two portrayed a very wide dispersion of ratings.

RESULTS

43. This section provides a visual presentation of the survey results. The results are interesting and we believe support the idea that rigorous development of a behaviorally anchored rating scale system be pursued.

IDEAL RESULTS

44. The results are those expected given (1) the abbreviated method used to develop the prototype, and (2) the first-time nature of its use. Again, we emphasize that the ideal results would be characterized by the commander-staff team being rated . . .

- in each sub-proficiency
- at exactly the same level of effectiveness
- by each participant.

Item No.	Partic. No.	Response						
		7	6	5	4	3	2	1
1.1	1			●				
1.1	2			●				
1.1	3			●				
1.1	4			●				
1.1	5			●				
1.1	6			●				
1.1	7			●				
1.1	8			●				
1.1	9			●				
1.1	10			●				
1.1	11			●				
1.1	12			●				
1.1	13			●				
1.1	14			●				
1.1	15			●				
1.1	16			●				
1.1	17			●				
1.1	18			●				
1.1	19			●				
				19				

Figure 6. Ideal set of Responses from All Observers of a Single Sub-proficiency

45. Figure 6 depicts ideal results for a sub-proficiency. The reason for this total agreement would be that the behavioral anchors for each sub-proficiency describe <u>perfectly</u> the behaviors observed by the command staff raters. Implied is that the raters agree that the relative effectiveness of the behavior in accomplishing tasks inherent in the proficiency is consistent with the numerical scale above the BARS.

46. This level of agreement is attainable only when the BARS have been developed using a methodology similar to that described originally in Smith and Kendall (1963), and refined successively by others in the years since their original work.

UNIT OF ACTION (UA) BATTLE COMMAND (BC) EXPERIMENT RESULTS

47. The UA BC Experiment results are presented in terms of the definitions of commander-staff rating variability described earlier.

Individual Rater Variability

48. Results for the first measure, individual rater variability, are shown in figure 7. The figure shows the range of rater responses across all sub-proficiencies by 19 participants. Each rater's total responses within each proficiency effectiveness level are shown in the bordered area. The "Variability" column shows the number of effectiveness levels assessed by each rater. The median shows the level of effectiveness that was exactly midway in the distribution of ratings from highest to lowest.

	Proficiency Scale									
Participant	7	6	5	4	3	2	1	Variability (# of cells)	Total Responses	Median
1	26			1				1+	27	7
2	4	2	6	8	3	1	4	7	28	4
3	2	5	11	7	5			5	30	5
4	6	11	5	2	1			5	25	6
5	1	4	12	5	1			5	23	5
6		6	7	16	3	2	1	6	35	4
7		5	14	7	7			4	33	5
8	1	14	9	4	2	2		6	32	5
9	2	5	1	11	10	2	4	7	35	4
10		1	6	20	7	1		5	35	4
11		8	4	11		1		4	24	4.5
12	2	6	5	12	1	2		6	28	4
13			4	20	4			3	28	4
14			7	5	12	11		4	35	3
15		5	18	12				3	35	5
16		9	13	13				3	35	5
17		3	18	9	5			4	35	5
18		1	2	9	3	6	4	6	25	3.5
19		2	9	22	1			4	34	4

Participant's most frequently reported proficiency scale
Participant's second most frequently reported proficiency scale
Participant's third most frequently reported proficiency scale

Figure 7. Individual Variability in Sub-proficiency Responses for All Participants

49. As the distributions of ratings in the figure shows, with the exception of one person, all other commander-staff raters demonstrated a reasonable degree of individual rating variability in terms of the different levels of effectiveness they assessed. Participant 1 assessed the effectiveness of the commander-staff team at level 7 in 26 of 27 sub-proficiencies, showing almost no individual rating variability in his assessments. We considered his assessments outliers compared to the results of the other participants.

Group Rating Variability Within a Single Sub-proficiency

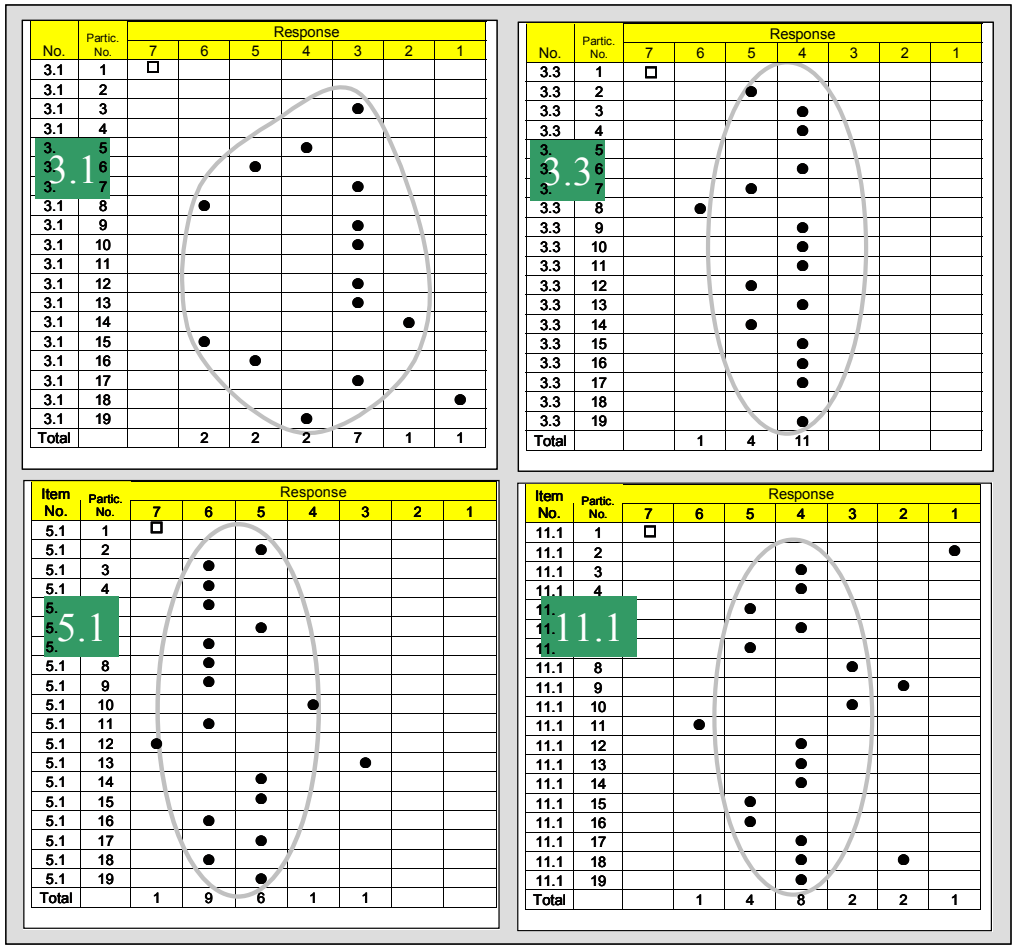

Figure 8. Group Variability Within a Single Sub-proficiency

50. Figure 8 shows the group ratings for four sub-proficiencies. This sample reflects the range of group ratings. The quadrants marked "3.1" and 11.1" show two sub-proficiencies where the participants' ratings were distributed across six levels of effectiveness. These strongly suggest that the sub-proficiency descriptions or the other factors contribute to the dispersion of assessments.[10] The sub-proficiency, "use doctrinal terms and standard formats," and the behavioral anchors are relatively unambiguous. Here, perhaps the combination of Objective Force concepts and Recognitional Planning Model rendered the sub-proficiency "not applicable."

51. The quadrants marked "3.3" and "5.1" represent the two sub-proficiencies with the narrowest distribution of assessments across the effectiveness levels. While neither approaches the ideal distribution in figure 6, "3.3" shows 11 of 16 ratings in cell 4 and a distribution of only 3 scales total. Quadrant "5.1" is not quite as tight, but 15 of 18 assessments are in two adjacent cells, and the outliers fall in either side of the center of

[10] The small square in the upper left corner of each quadrant indicates that Participant #1's ratings were discounted in the discussion of group rating variability.

mass. The distribution patterns for the remaining 31 sub-proficiencies are similar to or between the dispersed and narrow patterns in Figure 8 and are not shown here.

Group Rating Variability Across the Set of Sub-proficiencies

52. Figure 9 aggregates the distribution patterns in all 35 sub-proficiencies used during the experiment. The graphic contains considerable detail, but it is most useful looking at the global pattern of assessments.

Row	Prof. No.	Responses ~ All Participants 7	6	5	4	3	2	1	Variability (# of cells)	Total Responses	Median	At least 2/3's Responses fell within 1, 2, o r3 adjacent cells 1	2	3	Note
1	1.1	2	5	8	3		1		5	19	5		●		
2	1.2		1	6	7	2	2		5	18	4		●		
3	1.3			4	8	4	1		4	17	4		●		
4	1.4		1	6	7	2			4	16	4		●		
5	1.5	1	3	5	5	1	1		6	16	4			●	
6	3.1		2	2	2	7	1	1	6	15	3			●	
7	3.2		1	3	8	1	2		5	15	4		●		
8	3.3		1	4	11				3	16	4	●			
9	3.4			1	7	4	1		4	13	4		●		
10	3.5			3	6	3	1		4	13	4		●		
11	4.1			1	6	2	2		4	11	4		●		
12	4.2		3	5	5	2			4	15	5		●		
13	4.3			5	5		1	2	4	13	4		●		1
14	4.4		1	3	4	3	2		5	13	4		●		
15	5.1	1	9	6	1	1			5	18	6		●		
16	5.2		6	3	6	2	1		5	18	4.5		●		2
17	5.3	3	6	5	2	1			5	17	6			●	
18	6.1	1	2	7	5	3			5	18	5		●		
19	6.2	2	5	6	5				4	18	5			●	
20	6.3	3	7	5	2	2			5	19	6			●	
21	6.4	2	1	8	4	2			5	17	5		●		
22	6.5		4	2	6	2	1		5	15	4			●	
23	6.6	2	7	5	3				4	17	6		●		
24	6.7		3	3	8	2	1		5	17	4		●		
25	10.1		2	6	7	1			4	16	4.5		●		
26	10.2		1	5	6	3	1		5	16	4		●		
27	10.3		1	1	4	1	1	3	6	11	4				2
28	10.4			2	8	4	1	1	5	16	4		●		
29	10.5		1	2	6	2		2	5	13	4			●	
30	11.1		1	4	8	2	2	1	6	18	4		●		
31	11.2		2	8	5	3			4	18	5		●		
32	11.3			3	8	2	2	2	5	17	4			●	
33	11.4	1	3	5	7		1		5	17	5		●		
34	14.1	1	4	5	5		1	1	6	17	5			●	1
35	14.2	1	4	5	5	1	1		6	17	5			●	
												1	23	10	

Note: 1 Responses in two separated clusters

2 Responses continuous but in two lobes

No. of responses = or > 50% (1/2)

No. of responses = or > 50% (3/8)

No. of responses = or > 50% (1/4)

Figure 9. Group Variability Across the Set of Sub-Proficiencies

General Comments on the Distribution of Responses

53. The shaded "centers of mass" on the left side show variability among the group across all sub-proficiencies. While this variability was expected, interviews would have been helpful in determining how the participants actually understood and used the survey form. It would have been useful to ask (1) if the participants actually perceived differences in levels of effectiveness among the sub-proficiencies, (2) if the behavioral anchors actually captured behaviors the participants saw or could interpolate, or (3) if the group was simply applying the numerical scale values on the survey sheet to arrive at an innately sensed level of effectiveness.

54. The encouraging aspect of the variability analysis is that the pattern of responses for each sub-proficiency is clustered at one level of effectiveness or another. This suggests that validity questions not withstanding, most sub-proficiency descriptions and their behavioral anchors were "close enough" to establish a sense among the participant group of their collective effectiveness in the specific team tasks. This collective "sense" is reflected in the Median column.

55. Anticipating some degree of variability in the group ratings of each proficiency, the team determined that an acceptable level of agreement would be indicated by 66% of evaluations falling in two adjacent cells, or a single cell. The bullets down the right side show that in 22 of the 35 sub-proficiencies, 66% of the responses were in one or two adjacent cells.[11] Given (1) the prototype nature of the instrument, (2) the focus on the Objective Force staff structure, (3) the introduction of the Recognitional Planning Model, and (4) the abbreviated orientation of the participants to the BCP, 66% in 21 sub-proficiencies is promising as a first step "proof of principle."

56. The collective "sense" of the team's effectiveness in each sub-proficiency is reflected in the Median column. An assessment in the "4" column indicates the team "meets standards." "7" indicates the team "exceeds standards;" "1" indicates "below standards." The group sensed that it met standards or better in 34 of the 35 sub-proficiencies. It is difficult to assess this interesting, positive result. At least six variables could have resulted in a less sanguine total assessment: newness of the team; newness of the concepts, system capabilities, structure, and procedures; and lack of fidelity in the simulated C4ISR system. Conversely, the positive result could have been leveraged by three other variables: experienced personnel, entire team in one location, and tactical situations that were more familiar than expected given the 2010-2015 future time-frame. The groups' sense that its effectiveness was below a "4" occurred only once, in sub-proficiency 3.1. Not surprisingly, given the doctrinal situation, sub-proficiency 3.1 states "Use doctrinal terms and standard formats." Parenthetically, given the lack of detail in the Objective Force documents, an assessment of "4" or higher would have indicated a an apparent validity question.

Specific Comments on Each Proficiency

57. The following summary considers only those sub-proficiencies in which the group assessment of team proficiency achieved the 66% agreement criterion described earlier.

[11] Rows 16 and 27 are not included because the two cells containing 66% of the responses are not adjacent.

One – Clarify Expected Roles and Contributions of Individuals-Teams

58. The staff responded that the commander's and senior staff members' emphasis and involvement in the clarification of roles were well above the standard. They assessed the efforts of individuals to proactively learn their roles to be above standard, as well. They were more critical of their ability to support the decision-making process, reporting that they met the standard, but no more. The group sensed its efforts to purposefully acquire tacit knowledge and "know how" was above standard, as was the groups' willingness to contribute information and opinions during open planning sessions and to ask questions and elicit guidance at appropriate times. The ARL analyst observing the experiment concurred with the groups' overall assessment.

Three – Establish Effective Information Exchange Practices

59. The staff reported that they met standards in four of the five sub-proficiencies. They transferred information in a clear, timely and complete manner. Incoming information was generally verified. The staff generally assessed the information in terms of its implications for the on-going plan. They were attentive to acknowledging receipt of inbound information and following up to ensure they received acknowledgement for information they had sent. Use of doctrinal terms and standard formats is a component of this proficiency, but the group rating did not achieve the 66% criterion. The Objective Force concepts and structure and the RPM introduced many new terms. Given the time required to build individual vocabularies, this assessment is not surprising. In the normal rush of daily business, building a glossary of new terms is probably a low priority. Nonetheless, once the command group and staff is dispersed in six different vehicles, a common doctrinal language will take on even greater importance.

Four – Establish Supportive Behaviors and Error Monitoring

60. The key sub-proficiency in this series is anticipating the information needs of others and taking action to pass the information. The staff's effectiveness was above standard. The level of experience among the staff was very high, and each individual member, receiving a piece of information, probably has an innate understanding of which other persons need to have that information.

Five – Align Decision Authority with Decision-making Capacity

61. This proficiency focuses on the concepts of mission command and delegation of authority to subordinates to make specific decisions. The staff concluded that evidence of mission command was strong throughout the planning process, and continued through the execution phase. The staff also reported a high level of effectiveness in the delegation of authority to make certain specified decisions, but due to considerable variability, the group rating did not achieve the 66% criterion. Proficiency Five received the staff's highest overall assessment of effectiveness. The ARL analyst concurred that the concept of mission command was strongly evident in the commander's discussions and during execution.

Six – Employ Proper Mix of Decision Strategies for Each Situation

62. The staff assessed the command's adjustment of decision strategies almost as high as the Mission Command proficiency. The ARL analyst's observations were higher than the staff's. His observations, not included in the figures, are that the commander-staff

team was more effective in this proficiency than any other. The commander easily shifted between recognitional and analytical decision processes, depending on the clarity of the situation and available time. The staff adjusted to the commander's processes as easily, providing necessary information on cue and responding quickly to taskings.

Ten – Balance Push-Pull of Information to Decision-makers

63. The staff evaluated its effectiveness in staying focused on relevant and exceptional information as being above standard. The same assessment was made of the staff's understanding when to push information and when to pull it. Techniques for managing information overload were assessed as being lower, but still meeting the standard. The ARL team's sense was that the experiment conditions did not generate high volumes of information as has been seen in fully resourced field AWEs, and that the information push-pull requirements were exercised only moderately. This proficiency also includes the integration of liaison officers into the push-pull of information, and it includes the use of trusted officers as "directed telescopes." The lean manning levels of the UA brigade and battalion staffs did not permit the use of liaison officers or officers detailed as "directed telescopes." As a note for the future, although Change 1 to the Objective Force O&O lists the dispatch of liaison officers to higher and adjacent headquarters as a UA CIC responsibility, the cell organization graphics do not identify a specific team or vehicle assets to support this function. Recommend raising the question of resourcing liaison person and equipment to the Objective Force planning community.

Eleven – Maintain Attentional Scanning Across Multiple Decision Threads

64. This proficiency is the heart of the current operations monitoring function. The staff concluded it met standards with respect to individual staff members having been assigned specific monitoring duties. The staff was more effective in the actual monitoring function, and particularly in monitoring the commander's decision points. Finally, the staff considered they met standards for taking prompt action when they identified a variance or acquired other relevant or exceptional information from the monitoring activity.

Fourteen – Anticipate and Prepare for the Emergence of Complexity

65. Anticipation and preparedness for the emergence of complexity requires the commander-staff team to be exceptionally alert for the unusual to happen and alert to seemingly unrelated events. It also requires rapid response to confirm and act on these happenings. The medians for each sub-proficiency suggest the staff assessed its behaviors in this proficiency as above standard, but the variability was very large, and the evaluations indicate lack of agreement. Still, the median of five is worthy of comment because it may point to the use of the RPM "pre-mortem" in alerting the staff to unusual and unpredicted events. As no genuinely emergent events occurred during the scenarios, this assessment is probably a reflection of their being at least mentally alert for the emergence of complexity, if not actually tested in this important staff skill.

ARL Assessment

66. Setting aside the abbreviated development process and all other experiment specific variables, and in the context of the experiment, that is, a . . .

- new staff
- new staff procedures and techniques
- new Objective Force organizations and tasks
- unfamiliar Objective Force concepts
- unfamiliar C4ISR system
- unfamiliar recognitional planning model (RPM)
- compressed training
- but highly competent individuals able to make the best of the conditions

. . . the aggregate "center of mass" responses summarized above, look and feel "about right." The staff appeared to be able act effectively on all events and requirements that arose. The Objective Force Organizational & Operational Plan provides more staff members in the cells and vehicles than were used during the experiment. These additional persons, if they had been present, would have made the cells' work easier. Nonetheless, the staff still functioned effectively. For these reasons, and particularly because of the staff's own assessment of its effectiveness, we are confident stating the UA brigade headquarters, as represented during the experiment, demonstrated that the command and staff structure enables battle command.

PERSPECTIVES RELATIVE TO THE BCP

67. The content of the proficiencies and sub-proficiencies is a multi-level mix of doctrine, tactics, MTP tasks, and relevant insights into battle command behaviors from current research. The behavioral anchors provided by the military subject matter analysts include personal experiences, examples derived from the literature (Leedom), examples provided by others, and estimates of behaviors that are plausible.

68. The behavioral anchors demonstrate the range of actual critical incidents that could be generated by an efficiently resourced team of active duty subject matter experts.

69. The techniques for developing psychometrically valid and reliable BARS are straightforward. The process requires only a venue, normal office automation, a team of experienced practitioners (officers & senior NCOs), and a small team of analysts/facilitators. The officers and NCOs are critical to the process. They generate the list of behaviors and they validate each critical incident or behavior generated by the group.

70. Since the original BCP BARS concept was formulated, continued study has identified additional refinements to the analysis of battle command goals and proficiencies. Recent work provides greater granularity in the impact of C4ISR digitization on battle command proficiencies, and the need to highlight behaviors directly related to the commander-staff teams use of the human-computer interface and digitized operational displays. These additional insights would be available during a formal BARS development process. The insights would materially assist the officers and senior NCOs in articulating the critical incidents at a detailed level that are the basis of the subject matter experts' contributions to the development process.

71. The short-term cost of fully and formally developing the Battle Command Proficiency BARS for a modern C4ISR environment is small in comparison to the benefits projected by this first step proof of principle. A rigorously developed BCP BARS product would not only support continued C4ISR assessment, it would enhance unit training, readiness assessment, and unit initiated self-improvement and organizational learning, as well.

REFERENCES

Code of Best Practices for Experimentation. Washington, DC: Command & Control Research Program, DoD C3I. 2002

Cummings, L.L., and Donald P. Schwab. *Performance in Organizations: Determinants and Appraisal*. Glenview, IL: Scott, Foresman, 1973.

_____. *Development of a Battle Command Staff Proficiency Measurement System I*. Andover, MA: Dynamics Research Corporation, 2001. (Prepared under Contract DAAL001-95-C-0015 for the U.S. Army Research Laboratory, Aberdeen Proving Ground, MD 21005 and Headquarters, III Corps & Fort Hood, TX 76544).

Grubb, Gary; Simon, Robert; Leedom, Dennis K.; Zeller, J. *Development of Candidate Crew Coordination Evaluation Methods and Materials*. Wilmington, MA: Dynamics Research Corporation. 1992. (Prepared under Contract MDA903-92-D-0025 for the U.S. Army Research Institute for Behavioral and Social Sciences, Alexandria, VA 22333).

Leedom, Dennis K. *Development of a Framework for Assessing C4ISR Performance: Critical Battle Command Proficiencies*. (Working Paper) 2d Draft, May 26, 1999,U.S. Army Research Laboratory, Aberdeen Proving Ground, MD.

Murphy, James W. *Use of Behaviorally Anchored Rating Scales (BARS) to Complement the Management by Objectives (MBO) and Fitness Report Components of the Marine Corps Performance Evaluation System*. MMAS Thesis, USA Command and General Staff College, Fort Leavenworth, KS, 1980, 110 pages plus six appendices.

Smith, Patricia Cain, and L.M. Kendall. "Retranslation of Expectations: An Approach to the Construction of Unambiguous Anchors for Rating Scales." *Journal of Applied Psychology*, 47, No. 2 (April, 1963), 149-155.