REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)*
072814

2. REPORT TYPE
FInal report

3. DATES COVERED *(From - To)*
26Jun2013-26Jul2014

4. TITLE AND SUBTITLE
The Testing Phase for the Small unit Decision Making (SUDM) Assessment Battery: Final Report for Contract Option II

5a. CONTRACT NUMBER
N00014-12-C-0459

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Ross, Karol, G., PhD
Rivera, Iris, D., PhD
Lineberger, Regan, E., PhD
Vogel-Walcutt, Jennifer, PhD
Phillips, Jennifer, K.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Cognitive Performance Group of Florida, LLC.
3662 Avalon Park East Blvd., Suite 205
Orlando, FL 32828

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
875 North Randollph St.
Arlington, Va 22203-1995

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
A002

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The Small Unit Decision Making (SUDM) Assessment Battery is designed to measure the decision-making proficiency of infantry small unit leaders. The goal is to provide assessment of the cognitive readiness of Marine Corps maneuver squad leaders over time as experience and learning opportunities vary.
This report summarizes the testing phase of development. Participants consisted of Lieutenants in the Basic Officer Course at The Basic School and enlisted Instructor-Advisors working with that course. Some instruments in the battery had less than acceptable, consistent internal consistency reliability. Some instruments did not discriminate well among the participants as examination of histograms showed skewing to the right or left or truncated clusters indicating similar scores across the sample. Erratic responses to self-report instruments or skipping instruments were found for some instruments. Observation indicated test-taking fatigue, cognitive overload, and problems with the organization of the test materials as the causes. In the finalization phase, the data from all instruments except the cognitive flexibility scale will be subjected to further psychometric analysis to determine the construct and predictive validity of the battery. Results will determine which constructs underlying squad leader decision making can be meaningfully measured to assess overall proficiency. Items from instruments could be

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Karol G. Ross, PhD, Chief Scientist
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UU	16	**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U			407.282.4433

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

The Testing Phase for the Small Unit Decision Making (SUDM) Assessment Battery: Final Report for Contract Option II

Prepared by:

Karol G. Ross, Iris D. Rivera, Regan E. Lineberger, Jennifer Vogel-Walcutt, and Jennifer K. Phillips

Cognitive Performance Group
3662 Avalon Park East Blvd., Suite 205
Orlando, FL 32828

Cognitive Performance Group

Prepared for:

Office of Naval Research
875 North Randolph St.
Arlington VA 22203-1995

Contract Number: N00014-12-C-0459
28 July 2014

Approved for public release; distribution is unlimited

Abstract

In support of *Marine Corps Vision and Strategy (MCV&S) 2025 Implementation Planning Guidance* document Task 1 to develop a plan to improve small unit leader ability to assess, decide, and act in a more decentralized manner, the Marine Corps Training and Education Command (TECOM) created the Small Unit Decision Making (SUDM) initiative. One objective of the initiative is to develop a SUDM Assessment Battery to measure the decision-making proficiency of infantry small unit leaders over time. The purpose of this report is to summarize the testing phase conducted under Option II of the SUDM Assessment Battery contract. This phase of the research used a version of the battery based on a developmental model of maneuver squad leaders and on a multi-dimensional approach to determining decision-making proficiency. The purpose of the *testing phase* was to make final adjustments to the current instruments, administration, and scoring protocols, as needed, using a larger sample than the pilot group available during the development phase. The quality of the instruments based on use with the larger sample was determined, administration was examined, and the scales were examined for usefulness and meaning. In the *finalization phase* of this research, the results of that testing phase will provide the data for further psychometric analysis to examine the reliability and the construct and predictive validity of the battery. Results will determine which constructs underlying squad leader decision making can be meaningfully measured to assess overall decision-making proficiency and support insight into performance. During finalization, constructs will be combined if needed based on analysis, or scales could be reduced to their most meaningful items, and reduction in the size and structure of the battery will likely result. The final structure of the battery will be determined, and it will be packaged for use by non-researchers. Finally, the battery will be extended to a version for platoon commander assessment.

Contents

Tables

Executive Summary

Introduction

The *Marine Corps Vision and Strategy (MCV&S) 2025* calls for the Marine Corps to be the nation's expeditionary force of choice and to demonstrate the ability to rapidly deploy to a wide range of complex and irregular operating environments as lean, agile, and adaptable individuals and units. This vision is supported not only by changes to training, education, and experiences for small unit leaders, but also by creating better options to assess decision-making proficiency as a means of assessing the status of and improvements over time in *cognitive readiness*[1] across the force. The decision dilemmas faced by squad leaders are too numerous to count, let alone test as individual performance items. Furthermore, as is the case in cognitively complex performance environments, seldom if ever can a single best decision be identified for a given tactical problem. Prior assessment efforts have overcome these challenges by scoping the assessment space to a specific, well-defined set of performance parameters, or by relying on subject-matter expert (SME) ratings of decision quality as a means of quantifying decision performance. The multi-dimensionality of decision making is lost in the assessment process. These approaches also do not lend themselves to the Marine Corps' requirement for a scalable, generalizable assessment capability that predicts decision performance across a range of operational settings. Therefore, the Small Unit Decision Making (SUDM) Assessment Battery research project was undertaken to fill that gap.

Method

The period of performance for Option II was 26 June 2013 – 25 July 2014. Two tasks were required for that period—field testing and revision of the battery contents, administration, and scoring, as needed, prior to the Finalization Phase. During this period, data were collected by administering the battery at The Basic School (TBS) to the Basic Officer Course (BOC) companies completing the six-month course, both before and after the course. Participants consisted of a sample of Lieutenants (Lts) provided by TBS for each course beginning in FY14 and all the NCOs who were participating in each course to improve their ability to perform as TBS instructors as part of the Enlisted Instructor-Advisor Initiative (Desgrosseilliers & Hoffman, 2014). Only NCOs were assessed from two companies from the final FY13 courses as a pilot test of the battery in the TBS setting. In addition, ratings of NCO performance were obtained from supervisors and copies of the Command Evaluation Form—rating forms for each student completed by instructors—were also gathered. These two forms are performance criteria which will be analyzed in Option III to follow. In exchange for the opportunity to collect a large sample of data, our research team will offer insights into the impact of the Enlisted Instructor-Advisor Initiative in a separate project. The SUDM Assessment Battery project will use a portion of the data collected in FY13 and FY14 to support completion of the battery which is scheduled for 23

[1] "*Cognitive readiness* is the mental preparation (including skills, knowledge, abilities, motivations, and personal dispositions) an individual needs to establish and sustain competent performance in the complex and unpredictable environment of modern military operations," (Morrison and Fletcher, 2002, p. I-3).

September 2014. Data collection will continue at TBS through March 2015 to complete the impact analysis study under a separate contract. The final report on the Enlisted Instructor-Advisor Initiative will to be delivered in June 2015.

Findings

Quality of the instruments was measured by comparing reliability scores from the literature to those calculated based on the current data sets. While reliability and validity data were available in the literature for many of the instruments, our team re-analyzed internal consistency reliability for this population which is much larger than the earlier pilot study for this project and gave us insight into reliability for the military population. Results show that most of the instruments remained constant in their properties by maintaining similar reliabilities over time (similar at pre- and post-administration) and generally maintained the results found in the literature. Five of the ten instruments reviewed did not meet the reliability cutoff of .75 consistently. All will be retained for analysis in the finalization phase to determine if some items are of use for the final battery. Some instruments were skipped by some respondents and others were answered erratically. Observation and analysis indicate that the poor response behavior by some respondents is due to fatigue and cognitive overload.

Current findings show that the performance measures in the battery are significantly correlated though they were developed to assess different aspects of the decision making constructs. Specifically, the Decision Requirements Interview (DRI) correlated significantly with the SUDM SJT sensemaking subscale (r=0.153*, p=0.014, n=255), and the SUDM SJT correlated significantly with the Adaptability SJT (ASJT; r=0.193**, p=0.002, n=260), suggesting that these instruments vary similarly and measure associated constructs.

Finally, the DRI correlated significantly with years of service (r=0.140*, *p*=0.024, N=260), in that those with more years of service performed better on the DRI suggesting that it can distinguish between levels of expertise.

Recommendations

Based on findings from the testing phase, several changes are recommended for the battery. First, it is recommended that self-report measures be separated from the performance measures during administration by administering SJTs on a separate day from the booklet session to reduce fatigue. As long as a test booklet is used for administration, the back-to-back format may have to be replaced with single-sided pages to insure each page is considered by the respondent. If the administration is converted to a computer-based presentation at some point in the future, fatigue and cognitive overload can more easily be ameliorated. Second, we recommend the creation of a custom instrument, such as paired scenarios to test cognitive flexibility, i.e., to assess the ability of the respondent to transfer knowledge and principles from one setting to another in a flexible manner. Future analysis, such as a factor analysis, will provide better insight into the clusters of items that best explain the target constructs by examining the relationships among instruments and within the overall pool of items. Removing or adding items or instruments will be addressed as part of Option III as a result of the psychometric analysis.

Introduction

Requirement

Infantry small unit leaders represent one of the most critical positions on the modern battlefield. They form the tip of the spear against irregular threats in mission environments characterized by extreme levels of complexity. Operations in Iraq and Afghanistan have plainly demonstrated the broad decision-making responsibility of small unit leaders and the strategic failures that result from poor judgment. Future operations are likewise expected to require small unit leaders who can quickly recognize and adapt to evolving situations and make sound decisions that achieve the mission objectives while mitigating against negative second and third order effects.

The Marine Corps recognizes the vital role of the small unit leader. The stated vision of the *Marine Corps Vision and Strategy (MCV&S) 2025* is for the Marine Corps to be the nation's expeditionary force of choice and to demonstrate the ability to rapidly deploy to a wide range of complex and irregular operating environments as lean, agile, and adaptable individuals and units (U.S. Marine Corps, n.d.-a). In recognition of the small unit leader's role in that vision, one directive of the *MCV&S 2025 Implementation Planning Guidance* document (U.S. Marine Corps, n.d.-b) is to develop a plan to improve the small unit leader ability to assess, decide, and act in a more decentralized manner. Similarly, the Commandant's Planning Guidance (CPG) specifies a task to improve training and experience levels for maneuver unit squad leaders in support of decentralized operations in the 21st century hybrid threat environment (U.S. Marine Corps, 2010). In response to these demands, the Marine Corps Training and Education Command (TECOM) institutionalized a Small Unit Decision Making (SUDM) initiative. The goals of the SUDM program are not only to improve the training of decision-making skills across the population of noncommissioned officers (NCOs) who may serve as maneuver squad leaders, but also to measure individuals' decision-making abilities as a means of assessing the status of and improvements over time in cognitive readiness across the force.

The challenges associated with measuring decision-making performance are many. Tactical decision making at the small unit level is a broad and unwieldy concept that cannot be defined as a discrete cognitive activity. While the work of Klein (1989) describes the Recognition Primed Decision process (RPD) as the most widely used process—without training or conscious thought—in situations such as squad leader decision making during operations, a single cognitive process is not the gist of RPD. Instead, decision making involves a number of cognitive processes and access to a knowledge base. Therefore, assessing and improving decision making requires a multi-dimensional approach to performance.

The decision dilemmas faced by squad leaders are too numerous to count, let alone test as individual performance items. Furthermore, as is the case in cognitively complex performance environments, seldom if ever can a single best decision be identified for a given tactical problem. Prior assessment efforts have overcome these challenges by scoping the assessment space to a specific, well-defined set of performance parameters, or by relying on subject-matter expert (SME) ratings of decision quality as a means of quantifying decision performance. The multi-dimensionality of decision making is lost in the assessment process. These approaches also do not lend themselves to the Marine Corps' requirement for a scalable, generalizable assessment capability that predicts decision performance across a range of operational settings. Therefore, the SUDM Assessment Battery research project was undertaken to fill that gap.

What Does the Battery Assess?

The SUDM Assessment Battery was designed to assess skills, abilities, and characteristics that support small unit decision making and to assess performance on tactical decision-making tasks relevant to small unit leaders—Marine Corps maneuver squad leaders and platoon commanders. Performance assessment items were designed to relate performance to the specific skills and characteristics already identified so as to show their application in action through scenario-based decision making.

The skills, abilities, and characteristics were previously defined by the SUDM initiative. Constructs to be measured were derived from a series of workshops and surveys of Marine Corps leaders, Marine Corps SMEs, and leading researchers. The workshops produced five competencies and a number of suggested cognitive and relational skills (CARS) which were hypothesized to account for the most critical processes underlying maneuver squad leader decision performance (U.S. TECOM, 2011). From these findings, TECOM selected five competencies and 10 of the CARS for further study as the underlying basis for the generation of a decision-making assessment battery. The five cognitive competencies are sensemaking, problem solving, adaptability, metacognition, and attentional control. The ten CARS are perspective taking, analytical reasoning, anomaly detection, change detection, situational assessment, cognitive flexibility, ambiguity tolerance, resilience, self-regulation, and self-awareness. We added the overarching construct of decision making and developed performance tests with subscales that relate to one or more of the constructs.

The instruments selected or developed for the battery include those that measure *traits* (difficult to change; require long periods of time and/or targeted training and experience to change), *states* (change with knowledge and experience more easily than traits; trainable), and *performance* (domain- and situation-specific decision making that can change with knowledge and practice). The constructs and their associated measurement instruments are identified as measures of states, traits, or performance and listed below in Table 1. Those identified as states and traits are measured by self-report instruments with indirect questions that provide scores allowing insight into the relative degree of the state or trait. Two of the SUDM initiative constructs do not have assessment instruments (change detection and anomaly detection). No measures that were generalizable to assessing proficiency or applicable to measuring various sizes of groups could be identified. Some scores on the assessment can be expected to change as a result of knowledge and experiences more quickly than others. However, generally, changes in the scores from the battery occur over long periods of time as mastery matures, and change varies based on experiences that broaden the knowledge base of individuals, practice and reflection opportunities and support to reflect on learning, and the strength of the trait in the individual. While traits are difficult to change, the Marine Corps needs to be aware of the distribution of factors contributing to good decision making under stress to understand the cognitive readiness of the force.

Table 1. Constructs, Assessment Instruments, and Instrument Type

Constructs	Assessment Instrument(s)	Acronym	State, Trait, Performance
Problem Solving	Personal Problem Solving Inventory	PPSI	S
Metacognition	Metacognitive Awareness Inventory	MAWI	S
Attention Control	Neuro-Cognitive Assessment	NCA	T
Adaptability	Adaptive Force Scale Situational Judgment Test	ASJT	P
Sensemaking	SUDM Situational Judgment Test	SUDM SJT	P
Perspective Taking	Differences in Empathy Scale	DES	T
Analytical Reasoning	Metacognitive Activities Inventory	MAI	S
Anomaly Detection			
Resilience	Brief Resilience Scale	BRS	T
	Connor-Davidson Resilience Scale	CDRS	T
Change Detection			
Situational Assessment	SUDM Situational Judgment Test	SUDM SJT	P
Cognitive Flexibility	Youmans Cognitive Flexibility Assessment	YCFA	S
Ambiguity Tolerance	Multiple Stimulus Types Ambiguity Tolerance	MSTAT	T
Self-Regulation	Personal Problem Solving Scale	PPSS	S
Self-Awareness	Freiberg Mindfulness Inventory	FMI	S
Decision Making	Decision Requirements Interview	DRI	P
	SUDM Situational Judgment Test	SUDM SJT	P

Interestingly, Morrison and Fletcher (2002) hypothesized a similar set of 10 “components” as relevant to cognitive readiness and suggested that these be measured even though “some aspects of cognitive readiness are not amenable to training…” (p. III-1). Their components of cognitive readiness are (1) situation awareness, (2) memory, (3) transfer of training (ability to apply knowledge and skills in one context to another context), (4) metacognition, (5) automaticity (rapid responses that do not substantially impair other processes), (6) problem solving (situation analysis, understanding goals, and developing a course of action to achieve goals), (7) decision making (reviewing different plans of action, assessing the probable impact of each, selecting one, and committing resources to it), (8) mental flexibility and creativity, (9) leadership, and (10) emotion (devise and select courses of action under stress).

What the Battery Can Provide to the Marine Corps

The battery assesses the range of traits and cognitive processes that are involved in decision making performance and includes performance items designed to show performance of a number of those processes. Given the multi-dimensional nature of the battery construction, the results allow the Corps, at a high level, to see the overall proficiency of the group of maneuver squad leaders and prospective squad leaders at any given point in time. Both current performance and underlying cognitive constructs can be aggregated to paint a picture of strengths and needs for improvement that can be addressed at the service level in line with the *MCV&S 2025 Implementation Planning Guidance* task of "improving Small Unit Leader intuitive ability to assess, decide and act...." Therefore, the intended use of the battery is at the policy level to influence the training, education, and experiences of the maneuver squad leader and prospective squad leader and to assess the impact of such actions in the overall community or sub-communities no more regularly than once a year. For example, the current Squad Leader Development Program, recently initiated by the USMC, could take a sampling as a baseline and then check every few years to see if the program was improving those constructs and the level of mastery in the squad leaders, and if either of two tracks for development currently proposed produces more improvement than the other.

The battery is not designed as a training effectiveness evaluation tool for discrete events or training products given that it was designed to be sensitive at the level of overall development over extended periods of time. To determine the training needs or determine the impact of discrete training outside a complete developmental program for squad leaders, portions of the battery could possibly be used. If the whole battery were used in the context of devising or adjusting specific training for a specific set of individuals, the best use would be to interpret the data to understand strengths and weaknesses (psychological predisposition) to target or compensate for in training and feedback as well as understanding specific performance strengths and weaknesses.

Any and all administrations must be based on the development of an atmosphere that is conducive to the participants responding to the battery seriously and deliberately and based on administration in a manner that best reduces fatigue and cognitive load from test taking in order to obtain the most informative data.

Structure of the Project—Develop, Test, Finalize

The SUDM Assessment Battery project consists of three phases—*development, testing, and finalization*—to achieve a reliable and valid battery sufficient for understanding small unit tactical decision making. *This report details the testing phase* of research with a sample derived from The Basic School (TBS). A version of the battery derived from the development process in Option I was used across a number of companies that comprise the class cohorts at the TBS Basic Officer Course (BOC).

The purpose of the *testing phase* was to make final adjustments to the current instruments, administration, and scoring protocols, as needed, using a larger sample than the pilot group available during the development phase. The quality of the instruments based on use with the larger sample was determined, administration was examined, and the scales were examined for usefulness. Scoring will be re-examined during the finalization phase.

In the *finalization phase* of this research, the results of the testing phase will provide the data for further psychometric analysis to examine the reliability and the construct and predictive validity of the battery. Results will determine which constructs underlying squad leader decision making can be meaningfully measured to assess overall decision-making proficiency and support insight into performance and cognitive readiness. During finalization, constructs will be combined if needed based on analysis, or scales could be reduced to their most meaningful items, and reduction in the size and structure of the battery will likely result. The final structure of the battery will be determined, and it will be packaged for use by non-researchers. Finally, the battery will be extended to a version for platoon commander assessment.

Method

The period of performance for Option II was 26 June 2013 – 25 July 2014. Two tasks were required for that period—field testing and revision of the battery contents, administration, and scoring, as needed, prior to the Finalization Phase.

During this period, data were collected by administering the battery at TBS to the BOC companies completing the six-month course, both before and after the course. Participants consisted of a sample of Lieutenants (Lts) provided by TBS for each course beginning in FY14 and all the NCOs who were participating in each course to improve their ability to perform as TBS instructors as part of the Enlisted Instructor-Advisor Initiative (Desgrosseilliers & Hoffman, 2014). Only NCOs were assessed from two companies from the final FY13 courses as a pilot test of the battery in the TBS setting. In addition, ratings of NCO performance were obtained from supervisors and copies of the Command Evaluation Form—a rating form for each student completed by instructors—are also being gathered for each company. These two forms are performance criteria to be analyzed in Option III.

In exchange for the opportunity to collect a large sample of data, our research team will offer insights into the impact of the Enlisted Instructor-Advisor Initiative under a separate project. The SUDM Assessment Battery project will use a portion of the data collected in FY13 and FY14 to support completion of the battery which is scheduled for 23 September 2014. Data collection will continue at TBS through March 2015 to complete the impact analysis study under a separate contract. The final impact analysis report is to be delivered in June 2015.

Participants

During the Option II timeframe, 11 separate data collections were conducted. Data from seven of the nine BOC companies being made available to the team were collected prior to beginning and following participation in the BOC. Table 2 provides a description of the participants by company. Not all participant data below had been prepared and were available for this analysis. Selected portions of the data were used for different analyses.

Table 2. Participants from the Basic Officer Course at The Basic School

	Pre			Post		
Company*	NCO	LT	Total	NCO	LT	Total
FY 13 E	15	0	15	14	0	14
FY 13 F	12	0	12	9	0	9
FY 14 A	15	45	60	10	43	53
FY 14 B	14	32	46	12	29	41
FY 14 C	8	56	64	x	x	x
FY 14 D	4	59	63	x	x	x
FY 14 E	11	51	62	x	x	x
FY 14 F	x	x	x	x	x	x
FY 14 G	x	x	x	x	x	x
Total	79	243	322	45	72	117

* An x indicates these assessments of the FY 14 companies will not be complete during the SUDM Assessment Battery project prior to the early July 2014 cutoff date for data collection and therefore not available for analysis in the SUDM Assessment Battery project.

Materials

The SUDM Assessment Battery measures the competencies and CARS previously determined by TECOM to be supportive of decision-making proficiency. Each of these constructs and the associated assessment instrument or instruments are shown in Table 1 above. Additional materials consisted of a supervisor rating form developed by our team to rate NCO performance in the BOC. Our research team is also collecting the Command Evaluation Forms, a TBS product. The complete sets of rating forms have not been obtained at this time and are not analyzed here.

Procedure

All SUDM Assessment Battery administrations consisted of two parts: (1) a classroom session for the administration of the test booklet in a group setting, and (2) individual interview sessions for the Decision Requirements Interview (DRI). To reduce cognitive load on the participant, the classroom and interview sessions were typically conducted on separate days. During the classroom sessions the participants were allotted three hours to complete the test booklet. On average, participants completed the booklet in less than two hours. The interview sessions were allocated two hours to complete. Informed consent was obtained at the start of either the classroom session or the interview, whichever occurred first. At all administrations, TBS provided someone to speak to the participants about the importance of diligently completing the assessment. Not all participants, especially the NCOs, attended those informational sessions.

Analysis

The process to prepare the data from each data collection for analysis takes approximately three weeks to complete. Some factors that influence the completion time for this process are sample size, expertise of the interviewers, and time of data collection (pre or post). Instructor Ratings and Command Evaluation Forms are collected following BOC completion, therefore the data preparation process will take longer for post data collections. Depending on the expertise of the interviewer, recorded interviews are reviewed by a second interviewer to provide accurate ratings. After all data entry is completed accuracy checks are conducted to insure all data entry is correct. All data entry is completed by hand via Microsoft Excel and then exported into SPSS for data cleaning. During the data cleaning process in SPSS, all missing data, reverse coding, subscales, and composite scores are computed. Preliminary analyses (i.e., descriptives, histograms, reliability) are computed to insure no mistakes were made during the data preparation process. Data analyzed for this report include pre and post data from the FY13 E and F companies, pre and post data from the FY14 A and B companies, and pre data alone from the FY14 C and D companies. The exception to this sample is the Youmans Cognitive Flexibility Assessment (YCFA) which was used only during FY14 C and D pre-assessment and FY14 A post-assessment only as it was selected for exploration after the data collection process was under way.

Findings

Quality of the Instruments

Quality of the instruments was measured by assessing validity and reliability scores from the literature and comparing internal consistency reliability scores to those calculated based on the current data sets. While reliability and validity data were available for many of the instruments, our team re-analyzed reliability for this population which was much larger than the earlier pilot study of sergeants and gave us insight into quality for the military population. Results show that most of the instruments remained constant in their properties by maintaining similar reliabilities over time (similar at pre- and post-administration) and generally maintained the reliability results found in the literature. Table 3 indicates which instruments have undergone a successful validation study of various types (convergent, divergent, construct, and criterion), as well as the published reliabilities and the results of the Option II data.

Table 3. Reliability and Validity of the Instruments

Instrument	Prior Validity	Prior Reliability	Current Internal Consistency Reliability	
			Pre	Post
Brief Resilience Scale	Convergent; Divergent	0.8-0.91	0.82	0.82
Personal Problem Solving Inventory	Construct	0.90	0.82	0.82
Personal Problem Solving Scale	Divergent	0.81	0.77	0.89
Freiberg Mindfulness Inventory	Construct	0.86	0.48	0.42
Connor-Davidson Resilience Scale	Convergent; Divergent	0.89	0.89	0.89
Metacognitive Awareness Inventory	Construct	0.95	0.94	0.96
Neuro-Cognitive Assessment	Construct	0.98	0.95	0.96
Metacognitive Activities Inventory	Construct	-	0.71	0.80
Multiple Stimulus Types Ambiguity Tolerance	Criterion	0.86	0.86	0.86
Differences in Empathy Scale	N/A	-	0.72	0.80
SUDM Situational Judgment Test	N/A	-	0.44	0.57
Adaptive Force Scale Situational Judgment Test	N/A	-	0.37	0.50
Youmans Cognitive Flexibility Assessment	Unknown	-	-	-

Note: Red reliability findings are below the .75 cut-off established for reliability.

Using a cut-off score for reliability of 0.75, the majority of the measures reached adequate levels of reliability during pre- and post-BOC administrations. However, a few previously developed instruments did not, specifically the Metacognitive Activities Inventory (MAI) used to assess analytical reasoning, the Freiberg Mindfulness Inventory (FMI) used to assess self-awareness, and the Differences in Empathy Scale (DES) used to assess perspective taking. Two performance instruments—the Adaptive Force Scale Situational Judgment Test (ASJT) and the newly created Small Unit Decision Making Situational Judgment Test (SUDM SJT)—also did not reach the reliability cut off of 0.75.

For the current study, the FMI had a 0.48 and 0.42 for pre and post reliabilities. The original finding from the literature was much higher at 0.86. However, items had been removed that were not answered or specifically found to be objectionable through comments from earlier participants. The remaining seven items were used as the FMI instrument for this study. It is hypothesized that the low number of items used from the scale contributed to the low variability in scores and consequently also to the low reliability. Due to the smaller number of items in the instrument, internal consistency reliability is harder to establish. The pre reliability of the MAI and DES fell below the cut-off score at .71 and .72 respectively, while the post-administration reliability of these two scales exceeded the requirement, .80 for each. However, response behavior was erratic and the scales were sometimes skipped by participants. Variability was low and the majority of individuals scored in the low range. Because the post reliability score met criteria for this round and erratic response or skipping response altogether may not be a function of the instrument, the MAI and DES will be retained for now but reconsidered in the review for the final battery. The FMI will be retained so the data can be included in the psychometric analysis to be conducted in Option III.

Finally, the SUDM SJT and ASJT displayed low reliability. However, because this type of assessment is qualitative in nature for scoring, obtaining inter-rater reliability can be a challenge. Further, because response behavior, distribution, and variability were all normal, these performance-based assessments will be retained in their current form.

Response Behavior and Efficiency of Administration

Findings about administration and efficiency of using certain scales include response patterns and an assessment of the instruments based on the resulting distribution of scores (skewed to right or left and variability of scores within the group). The Youmans Cognitive Flexibility Assessment (YCFA) is also reviewed. The YCFA was used in a limited number of administrations to ascertain if it would be of use in the battery.

Response behavior was not good on some of the scales. Participants skipped the DES often, possibly because it was a one-page instrument on the last page of a booklet printed front-to-back, though three other instruments were also sometimes skipped, the Brief Resilience Scale (BRS), the Neuro-Cognitive Assessment (NCA), and the MAI. It appears that participants started to show signs of test-taking fatigue once they reached the middle of the booklet, resulting in the skipped scales. Thus, because the Metacognitive Awareness Inventory (MAWI), NCA, MAI, Multiple Stimulus Types Ambiguity Tolerance (MSTAT), and BRS are in the second half of the test booklet and after the Situational Judgment Tests (SJTs), participants may have been tired or experiencing cognitive overload after making multiple decisions and therefore skipped or responded erratically to get through the book more quickly. Also, the BRS and DES are short surveys at either the beginning or at the end of the book and were more frequently skipped. Observation suggests this may have been done by accident. The front-to-back printing format may contribute to participants skipping these first and last instruments. Finally, the most frequently skipped or irregular responses were the NCA, DES, and MAI. Table 4 below summarizes the response behavior just discussed, as well as the distribution and the variability observed when histograms of each instrument were examined.

Table 4. Response Behavior, Distribution of Responses, and Variability of Responses

Instrument	Response Behavior	Distribution	Variability
Brief Resilience Scale[3]	Skipped[1]	Left - Skewed	Normal
Personal Problem Solving Inventory	Appropriate	Truncated	Low
Personal Problem Solving Scale	Appropriate	Truncated	Low
Freiberg Mindfulness Inventory	Appropriate	Normal	Normal
Connor-Davidson Resilience Scale[3]	Appropriate	Left-Skewed	Low
Metacognitive Awareness Inventory[3]	Erratic[2]	Slightly Left-Skewed	Normal
Neuro-Cognitive Assessment	Skipped[1] Erratic[2]	Plateau	Wide
Metacognitive Activities Inventory	Skipped[1] Erratic[2]	Truncated	Low
Multiple Stimulus Types Ambiguity Tolerance	Erratic[2]	Normal	Normal
Differences in Empathy Scale	Skipped[1] Erratic[2]	Right-Skewed	Low
SUDM Situational Judgment Test[3]	Appropriate	Truncated, Left-Skewed	Low
Adaptive Force Scale Situational Judgment Test[3]	Appropriate	Truncated, Left-Skewed	Low
Youmans Cognitive Flexibility Assessment[3]	Positive responses	Right-Skewed	Normal

Note: 1=skipped possibly due to booklet placement; 2=Erratic responses possibly due to fatigue; 3= high scores

A distribution is said to be skewed when the data points cluster toward one side of the scale more so than the other, creating a curve that is not symmetrical. In other words, the right and the left side of the distribution are shaped differently from each other. There are two types of skewed distributions. A distribution is positively skewed if the scores fall toward the lower side of the scale and there are very few higher scores. Positively skewed data is also referred to as "skewed to the right" because that is the direction of the long tail end of the chart. Therefore, *skewed right means there is a cluster of low scores*. *Skewed left* is negatively skewed which *means there are very few low scores and the scores cluster in the high end of the ranges*. Truncated denotes a tight cluster around the mean. This looks like a normal distribution but with the tails missing. Plateau denotes a wide, flat distribution.

Examination of the histogram distribution for each instrument revealed the following observations. Table 4 summarizes that five of the instruments (BRS, CDRS, MAWI, SUDM SJT, and ASJT) are left-skewed, i.e., the scores cluster around the higher end of the range. For the self-report measures (BRS, CDRS, and MAWI) participants completed the self-report questions in a manner that yielded a high score on those constructs (resilience and metacognition). For the performance measures (SJTs), participants scored medium to high on the test. The YCFA is right-skewed; however, low scores on YCFA are indicators of fast responses which is the desired outcome. This means that six instruments or about half of all instruments lead to high scores among participants. This is also reflected under the variability column, which demonstrates that about half of the instruments have low variability among participants. These instruments may not be sensitive enough to differentiate among participants and this can be due to the heavy reliance on self-report measures and a need for participants to choose socially desirable answers. Alternatively, this population might have a high level of resilience and metacognition. The other half of the instruments appears to have a normal distribution and variability. Future factor analysis and regression analysis may reveal that an efficient battery will only require a subset of the current instruments if they cover the most meaningful, useful constructs, have normal variability and do not result in skewed responses while still relating well to the performance measures. Conversely, new measures of some constructs may be needed to cover constructs important to TECOM. Low variability and high scores on the SJTs may indicate they are too easy.

The cognitive flexibility test (YCFA) was added as an exploratory performance scale after this testing phase had begun. To determine relevance, several actions were conducted. Specifically, correlations between the YCFA and the ambiguity tolerance scale, SUDM SJT, and ASJT, differences between pre- and post-testing, examination of the YCFA literature, and the definition used by the test compared to the one used by the Mastery Model were considered.

Upon closer examination of the literature, it was found that the YCFA is typically administered to one individual in a manner that allows close monitoring of responses between one item to another as the respondent decides how to move from one image to another while meeting the rules given. Data are collected on each separate decision. This is not practical in the battery administration as it is now conducted. Our administrations can take into account only the gross-level inverse relationship between high cognitive flexibility and the overall time it takes the participant to finish the test. Overall accuracy and differences in decisions within test performance cannot be considered. The Mastery Model definition of cognitive flexibility is, "Applying knowledge and principles of tactics and leadership differentially based upon the unique demands of the situation. Applying knowledge learned in one context to multiple relevant contexts." It is questionable whether this puzzle matches the Mastery Model definition since neither tactical skills nor leadership is required to solve the puzzle. Analyses of the YCFA were not significant.

Overall, our analysis suggests that the cognitive flexibility scale may not be sensitive enough to capture differences after exposure to interventions, it cannot be scored to the detailed extent represented in the literature, and it is not related to our definition of performance. Consequently, it is recommended that the YCFA be removed from the final battery. Since we have not identified another measure of cognitive flexibility that meets the criteria for inclusion in the battery, that construct would not be addressed.

Meaningfulness of the Testing Phase Data

As noted above, the constructs examined are similar to the underlying components of cognitive readiness identified through literature review by Morrison and Fletcher (2002). This relationship indicates that the constructs identified independently by the TECOM workshops prior to this project are likely to be important to understanding the cognitive readiness of small unit decision makers. The difference in our interpretation versus that of Morrison and Fletcher is that the constructs studied here are seen as parts of the multi-dimensional nature of decision making that operate in concert as decisions are made. Meaningfulness of the data will be more fully addressed in the Option III Final Report at the end of the project. At that time, the full data set will be ready for examination to determine the statistical relationships among our decision-making performance instruments and scores on self-report instruments, and between our instruments and performance ratings of NCOs and Lts.

Current findings show that the performance measures in the battery are significantly correlated though they were developed to assess different aspects of the decision making constructs. Specifically, the DRI correlated significantly with the SUDM SJT sensemaking subscale (r=0.153*, p=0.014, n=255), and the SUDM SJT correlated significantly with the ASJT (r=0.193**, p=0.002, n=260), suggesting that these instruments vary similarly and measure associated constructs.

Finally, the DRI correlated significantly with years of service (r=0.140*, p=0.024, N=260), in that those with more years of service performed better on the DRI suggesting that it can distinguish between levels of expertise.

Discussion

The Challenge of Assessing Small Unit Decision Making

Decision making is an attractive construct to address in the research community, because the essence of what happens in a tactical environment is dependent on decision making. Most of us feel we would know decision making when we see it, but upon closer examination, the complexity of the process does not lend itself to a consistently agreed upon definition. Previous attempts to measure decision making have approached it as a singular construct, thus instruments developed from that theoretical basis tend to lack the sensitivity required to distinguish all the cognitive processes that are exercised when making decisions. To measure decision making, we cannot examine only the act of comparing options, study the Marine's analysis of the constraints and benefits to committing resources in a particular way in the context of a set of goals, or measure the outcomes of carrying out a plan. Instead, our approach to understanding and assessing decision making is dependent on the assertion that the decision making that matters in today's hybrid warfighting environment is multi-dimensional and the different cognitive dimensions that work together during decision making can be assessed and supported to improve decision making.

Inherently, good assessment of decision making is time consuming. Subject-matter experts need extensive time and observation to understand and assess proficiency. To improve the assessment capabilities of the Marine Corps we must produce a concise battery that is minimally time-consuming but still informative, that can be easily administered, scored, and interpreted by non-researchers, and that does not place a heavy burden on the participants causing them to provide data that is not optimally useful. However, the battery must still take into account and measure the multiple dimensions of decision making and avoid reducing complex performance to that which is easiest to measure.

General Recommendations Derived from the Testing Phase

Based on findings from the testing phase, several changes are recommended for the battery. First, it is recommended that self-report measures be separated from the performance measures, administering SJTs on a separate day from the booklet session to reduce fatigue. As long as a test booklet is used for administration, the back-to-back format may have to be replaced with single-sided pages to insure each page is considered by the respondent. If the administration is converted to a computer-based presentation at some point in the future, fatigue and cognitive overload can more easily be ameliorated. Second, in place of the YCFA, we recommend the creation of a custom instrument, such as paired scenarios to test cognitive flexibility to assess the ability of the respondent to transfer knowledge and principles from one setting to another in a flexible manner. Future analysis, such as a factor analysis, will provide better insight into the clusters of items that best explain the target constructs by examining the relationships among instruments and the overall pool of items. Removing other items or instruments will be addressed as part of Option III once the psychometric analysis has been completed.

Finalization of the SUDM Assessment Battery

The tasks for Option III, the final part of the project consist of a psychometric analysis and extension of the battery to another domain. The psychometric analysis task includes finalization of the battery. Deliverables consist of an interim report (23 August 2014), a final version of the battery with an administration manual (23 September 2014), a final version of the battery that has been extended to a new domain (platoon commander version; 23 September 2014), and a Final Report (23 September 2014).

In the finalization phase of this research, the results of the testing phase will provide the data for further psychometric analysis to examine the reliability, the construct structure, and the validity of the battery. A number of research questions guide the finalization phases:

- Can the multi-dimensional nature of decision making be demonstrated statistically in terms of the contributions of different constructs to proficiency as measured by the use of performance-based measures and instructor ratings as criteria?
- Has the efficiency of the battery improved in terms of using the fewest constructs shown to be good predictors of performance administered in as little time as needed to show useful results?
- Is the battery valid? Do scores correlate with SME judgment?
- Are the Key Performance (KPA) scores (based on the Mastery Model and comprised of different combinations of instrument scores in the battery) indicative of overall level of mastery in terms of their relationship to the demographics and to other measures of performance?

Results of the testing will determine which constructs underlying small unit decision making are most meaningfully measured to assess overall decision-making proficiency and support insight into performance. During finalization, constructs may be reduced and refined to make the battery as effective and efficient as possible based on multiple regression analysis and factor analysis. The final structure of the battery will be determined, and it will be packaged for use by non-researchers. Finally, the battery will be extended to another finalized version of the battery for platoon commander assessment.

Limitations of the Testing Phase

Data from the testing phase had not been completely prepared for analysis at the time of this report due to the need to collect as much data as possible for the finalization phase during the testing phase. Priority of effort was to coordinate and conduct battery administrations. Analysis presented here is based on a subset of the total data that will eventually be available to finalize the battery.

The sample assessed during the testing phase is a convenience sample. Findings may be limited due to a large part of the sample consisting of Lts who are little more than new college graduates when starting BOC. This large portion of the sample is likely not even at the novice stage, but are rank beginners during the pre-assessment because they have almost no military knowledge and experience. The Lts may only achieve the level of advanced beginner as tactical decision makers by the time they finish BOC and complete the post-assessment. Likewise, many of the NCOs are not trained and experienced in tactical decision making given their military specialties, even though they have much more military experience than the Lts. Given this relatively novice sample, findings may be limited as to how well the research team can tease apart the relationships among constructs posited to exist in the population of maneuver squad leaders and prospective squad leaders when comparing self-report construct scores with performance scores.

Future Research and Development of the Battery

It is recommended that at the conclusion of the project, the final battery be implemented with a relatively large sample size from the desired population—prospective and current maneuver squad leaders. Prior to the end of the project, our team will seek to determine how many members of the data set have prior military experience in the infantry specialties in the NCO and Lt samples. Examination of this small group of more experienced infantry Marines more similar to the target population could yield more insight into how the battery constructs are related. Additionally, we recommend a sample for test-retest reliability be accessed after the battery is finalized to add another dimension of information about the quality to the battery.

Future research can also concentrate on constructs not addressed in the current battery—change detection, anomaly detection, and cognitive flexibility. It is likely that these constructs must be addressed with custom designed instruments that meet the criteria of the construct definitions derived during early work in the project. It is possible that more of the constructs currently measured by self-report should also be integrated into performance measures to create a better picture of the individual respondent or respondent group. As noted above, some current instruments may not address the construct sufficiently due to skewed results and low variability in scores. These instruments may need to be replaced to adequately cover the constructs of interest and given that no other appropriate scales have been identified, conversion to custom made performance items could be the most informative type of assessment.

Additionally, the battery should be converted into a computer-administered version following this project to mitigate test fatigue and cognitive overload by allowing the respondents to save their work, stop, and return when refreshed to a password protected assessment that must be completed within an adequate, designated amount of time from first login.

References

Desgrosseilliers, T. S., & Hoffman, R. (2014, March). The Basic School. *Marine Corps Gazette*, *98*(3), 8-20. Quantico, VA: Marine Corps Association and Foundation.

Klein, G.A. (1989). Recognition-primed decisions. In W.B. Rouse (Ed.), *Advances in man-machine systems research* (pp. 47–92). Greenwich, CT: JAI Press.

Morrison, J. E., & Fletcher, J. D. (2002). Cognitive readiness. (IDA Paper P-3735). Alexandria, VA: Institute for Defense Analysis.

U.S. Marine Corps [USMC]. (n.d.-a). *Marine Corps Vision & Strategy 2025*. Retrieved from http://www.onr.navy.mil/~/media/Files/About%20ONR/usmc_vision_strategy_2025_0809.ashx.

U.S. Marine Corps [USMC]. (n.d.-b). *Marine Corps Vision and Strategy 2025: Implementation Planning Guidance*. Quantico, VA: Author.

U.S. Marine Corps [USMC]. (2010). *The 35th Commandant of the Marine Corps Commandant Planning Guidance*. Washington, D. C.: U.S. Department of the Navy.

U.S. Marine Corps Training and Education Command [TECOM]. (2011). *U.S. Marine Corps small unit decision making workshop: Version 4.0* (Final Technical Report). Quantico, VA: Author.