Assessing Trainee Performance in Field and Simulation-based Training: Development and Pilot Study Results

Emily E. Wiese[1], Capt Ryan Nungesser[2], Ryan Marceau[1], Matthew Puglisi[1], Bill Frost[3]

[1]Aptima, Inc.
Woburn, MA
ewiese@aptima.com, rmarceau@aptima.com
mpuglisi@aptima.com

[2]United States Marine Corps
Camp Lejeune, NC
ryan.nungesser@usmc.com

[3]Cubic Corporation
Camp Lejeune, NC
bill.frost@usmc.com

ABSTRACT

Motorized Patrol Operations (MPO) provide Marine Corps commanders with effective means to watch and protect large areas of operations, as well as the ability to quickly adapt to different mission requirements. Marine leaders conducting a patrol must be vigilant and, above all, well trained, as constant changes in enemy tactics and an increasing need for motorized patrols require MPO leaders to make quick and effective decisions without command guidance. The actions taken by an MPO leader are dictated by split-second analysis and rapid decision making, which are difficult skills to teach and assess. Pre-deployment training takes a crawl-walk-run approach to teaching these skills, frequently involving a combination of classroom instruction, live field exercises and increasingly, simulated exercises.

Effective performance assessment and feedback are important to both live and simulated training. Without appropriate performance measures, it is difficult to systematically assess the readiness of trainees and the effectiveness of the training curriculum. Furthermore, without any guiding instructional framework, feedback given to trainees is dependent on the style of the individual instructors. Current methods of performance assessment for MPO still are largely informal in nature, relying on the abilities of instructor/observers to accurately remember or take notes on key aspects of performance. This can result in excluding critical aspects of performance and instructor bias in interpreting the results of the training exercise.

In this paper, we describe a recent effort to develop observable performance measures that assess MPO leader performance during both live and simulated training, using the COMPASS Methodology. We also describe the results from a pilot study in which we assess the utility and usability of these performance measures in both simulated and live exercises, as implemented in an observer-based assessment tool. We conclude with a discussion of potential implications of our results and future directions this research may take.

ABOUT THE AUTHORS

Ms. Emily Wiese is a Human Factors Engineer and product manager at Aptima, Inc. Her work involves developing performance measures for training in various military domains, including the AF AOC, Marine Motorized Patrol Operations, and Navy helicopter crews. Ms. Wiese develops tools to support observer-controllers take measures during training exercises and tools to support instructors create and use system-based measures. Additionally, she has developed an instructional program for US Navy Special Warfare Combatantcraft Crewman (SWCC) high-speed boat helmsmen, led an effort to develop a training system for the Military Decision Making Process (MDMP), and led an effort to develop training for Air Weapons Controllers. She received a B.S.E in 2001 and an M.S. in 2003, both in Industrial Engineering, from the University of Iowa.

Capt Ryan Nungesser is currently serving as the Operations Officer for the Advanced Infantry Training Company (AITC), School Of Infantry (East) aboard Camp Geiger, NC. Capt Nungesser has been in the Marine Corps since 2003, and has served as both a Rifle and Weapons Platoon Commander during two deployments to Iraq in support of Operation Iraqi Freedom. In addition to his current duties as the Operations Officer for AITC, Capt Nungesser

also serves as the Officer in Charge of the Infantry Squad Leader Course, Infantry Anti-Tank and Assault Leader Course, Infantry Mortar Leader Course, and Infantry Machinegun Leader Course. Capt Nungesser's personal decorations include the combat action ribbon, and a Navy and Marine Corps Achievement Medal with Combat Distinguishing Device. Capt Nungesser holds a B.S. in Human and Organizational Development from Vanderbilt University in Nashville, TN and is currently working on a Masters Business Degree from Boston University.

Ryan Marceau is an Industrial/Organizational Analyst at Aptima, Inc. He has experience in quantitative analysis, competency modeling, survey development, and training design. Mr. Marceau's interests are in applying principles from IO Psychology to the understanding and improvement of complex systems. He supports efforts including the identification of training requirements and associated training gaps for Air Force platforms and organizations ranging from Bombers and Jet Fighters to Information Operations teams in Air Operations Center. Mr. Marceau received a M.A. in Industrial and Organizational Psychology from Hofstra University, and a B.A. in Psychology, with a Minor in Business, from Saint Michael's College.

Matthew Puglisi is a Senior Military Analyst with over 18 years of service as a Marine artillery officer in the active and reserve forces, including combat and command experiences. At Aptima, he applies his knowledge of these domains to the development of military-related experiments by designing data collection methodologies, conducting data collector training, and performing data analysis, and also serves as the Company's Director of Business Development. During his military career, Mr. Puglisi served as an artillery forward observer for the First Battalion Eighth Marines during Operation Desert Storm and commanded Battery G, Third Battalion, Fourteenth Marines. Mr. Puglisi's personal decorations include the Navy Commendation Medal, the Navy Achievement Medal and the Combat Action Ribbon. Mr. Puglisi holds a Masters in Public Policy from Georgetown University, and a B.A. in Political Science from Siena College.

Bill Frost is a retired SSGT with the USMC, currently working for Cubic Corporation. Mr. Frost is key in developing and maintaining the simulation facility at the School of Infantry's Advanced Infantry Training Company.

Assessing Trainee Performance in Field and Simulation-based Training: Development and Pilot Study Results

Emily E. Wiese[1], Capt Ryan Nungesser[2], Ryan Marceau[1], Matthew Puglisi[1], Bill Frost[3]

[1]Aptima, Inc.
Woburn, MA
ewiese@aptima.com, rmarceau@aptima.com
mpuglisi@aptima.com

[2]United States Marine Corps
Camp Lejeune, NC
ryan.nungesser@usmc.com

[3]Cubic Corporation
Camp Lejeune, NC
bill.frost@usmc.com

INTRODUCTION

Motorized Patrol Operations (MPO) provide Marine Corps commanders with effective means to watch and protect large areas of operations, as well as the ability to quickly adapt to different mission requirements. Marine leaders conducting a patrol must be vigilant and, above all, well trained, as constant changes in enemy tactics and an increasing need for motorized patrols require MPO leaders to make quick and effective decisions without command guidance (Dieckhaus, 2006).

The actions taken by an MPO leader are dictated by split-second analysis and rapid decision making. Frequent and effective training before and during a deployment is critical to ensure that our Marine leaders are equipped with the necessary decision making skills to meet the challenges of today's battlefield. Unfortunately, teaching these types of decision making skills is difficult (Phillips, Shafer, Ross, Cox, & Shadrick, 2005). Pre-deployment training takes a crawl-walk-run approach to teaching these skills, frequently involving a combination of classroom instruction, live field exercises, and, increasingly, simulated exercises. Simulation-based training has been developed in part to reduce some costs of field exercises, while also providing a training environment that is readily available and easily configurable for different mission types. These environments allow pre-deployment practice in necessary critical thinking skills, a training strategy that has shown to improve decision making skills (Phillips et al., 2005). While simulation-based training cannot replicate the field environment completely, developers are increasingly creating systems to augment field exercise training, particularly for Marines (Lyssy & Munden, 2006; O'Bea, Bell, & Crabtree, 2006; Roby & Pedersen, 2006).

Training for MPO at the School of Infantry (East), hereafter SOI-E, employs both field exercises and simulation-based training. In this paper, we will describe a recent effort to develop observable performance measures that assess MPO leader performance during both live and simulated training. We will also describe the results from a pilot study in which we assess the utility and usability of these performance measures. Finally, we will conclude with a discussion of the implications of our results and future directions this research may take.

Performance Assessment Challenges in Live and Simulated Training

Effective performance assessment and feedback are important to both live and simulated training. Without appropriate performance measures, it is difficult to systematically assess the readiness of trainees and the effectiveness of the training curriculum (Gagne, Briggs, & Wager, 1992). Furthermore, without any guiding instructional framework, feedback given to trainees will be dependent on the style of the individual instructors. Current methods of performance assessment for MPO still are largely informal in nature, relying on the abilities of instructor/observers to accurately remember or take notes on key aspects of performance. This can result in excluding critical aspects of performance and instructor bias in interpreting the results of the training exercise (Biddle, Keller, Pitz, & Nixon, 2005).

The Focus

The primary goal of the current MPO measurement effort was to develop a comprehensive set of observer-based performance measures for Marine MPO and assess their utility and usability. To that end, we had several specific practical and theoretical goals in developing the measures.

1. **Training Environment**. In order to be truly useful in assessing trainee performance throughout their time at SOI-E, the performance measures had to be useful and appropriate in both field and simulated exercises.
2. **Measure granularity**. The performance measures developed needed to focus on the decision making capabilities of the MPO leader during a variety of MPO scenarios. Therefore, they could not be so general as to provide no meaningful insight into specific decisions made during a training session. Similarly, the measures could not be so detailed (and, consequently, numerous) so as to overwhelm the instructor making the ratings. Finding the appropriate level of granularity with regards to performance assessment was critical.
3. **Inter-rater Reliability.** While logistics precluded a formal assessment of the inter-rater reliability of these measures, we were committed to developing measures that would be interpreted similarly by multiple instructors. This is critical from both theoretical and practical perspectives; if these data are to be used in making assessments of the operational readiness of trainees, it is imperative that all trainees are rated consistently, with minimal instructor bias. From the standpoint of practicality, the instructors will periodically rotate out of SOI-E. Therefore, it must be clear to new instructors how to interpret and rate the measures with little formal training.
4. **Ease of Use**. The performance measures and corresponding rating instrument had to be very easy to use in order to facilitate widespread adoption of the measures by the instructors.

MEASURE DEVELOPMENT APPROACH

Our approach to developing effective and reliable observer-based measures employed the **CO**mpetency-based **M**easures for **P**erformance **AS**sessment **S**ystems (COMPASS) Methodology, described in detail in an article by MacMillan et al. (MacMillan, Entin, Morley, & Bennett Jr., In press). While subsequent COMPASS efforts have resulted in some overall refinement and domain-dependent variations in implementation, the underlying theory and protocol are the same as those described in the article.

At the heart of the COMPASS methodology is the insight that individual and team performance measures must be developed using a combination of subject matter expertise and psychometric theory. Subject matter expertise provides the operational knowledge needed to create measures that are domain-relevant. Psychometric theory ensures that the measurement instruments and procedures are valid and reliable. Furthermore, performance measurement that is based on clearly defined, behavioral objectives allows for unambiguous assessment of both trainee learning and training effectiveness (Gagne et al., 1992).

Subject matter experts (SMEs) are critical to the successful execution of the methodology. Over the course of our measure development effort, we held three knowledge elicitation sessions with SMEs from SOI-E. The first and third sessions were group interviews; the second session was composed of individual or small group interviews. Five SMEs were involved in the first session, ranging in rank from Sergeant to Captain. In the second session, we interviewed 12 SMEs, ranging in rank from Sergeant to Major. The third and final session involved seven SMEs, ranging in rank from Sergeant to Gunnery Sergeant. All of the SMEs were instructors at SOI-E.

Identify Performance Indicators via Mission Deconstruction

One of the goals of the COMPASS Methodology is to develop measures that directly assess MPO training objectives, thus ensuring that MPO leaders have the knowledge and skills necessary to be mission ready in an operational environment. To achieve this for MPO leaders, we first deconstructed the phases of the MPO mission to identify *performance indicators* (PIs) during the first knowledge elicitation session. For our purpose, we have defined PIs as high-level observable behaviors that an expert would use to recognize whether an individual or team is performing well. They are *tangible performance objectives* necessary for being an expert MPO leader, defined at a level that supports subsequent measure development.

Table 1. Example mission phase and example performance indicators identified for that phase.

Mission Phase	Performance Indicators
Conduct (primary): React to Contact (secondary)	Properly communicate contact report
	Employ appropriate escalation of force

We first broke down the MPO mission into phases and sub-phases. From three primary and 19 secondary mission phases identified, we generated 47 PIs. Table 1 shows an example primary and secondary mission phase and associated performance indicators.

Develop Performance Measures based on Performance Indicators

After the PIs were identified, we gathered information with which to develop specific *performance measures* (PMs) associated with each PI. During the individual/small group interviews, we worked closely with the SMEs to identify desirable and undesirable behaviors associated with each PI. Conducting this phase of the data collection on an individual basis is critical as it allows us to obtain a variety of perspectives about what behaviors constitute good and poor performance for each PI. In this manner, all of the SMEs' interpretations and perspectives on the PIs identified in the first session are heard. At the end of the individual/small group interviews, we had accumulated a large body of data to describe the observable behaviors critical in assessing behavior during motorized patrol operations.

We then used this data to develop a comprehensive set of observable performance measures for assessing the performance of the MPO leader. Each performance measure is composed of a question and a rating mechanism. Whenever possible, five-point Likert scales with behavioral anchors at points one, three, and five were used as the rating mechanism. For all Likert scale performance measures, a '1' indicates novice performance, a '5' indicates expert performance, and '3' indicates performance by someone who is neither an expert nor a novice, but demonstrates average performance. In addition to behaviorally-anchored Likert scale questions, "Yes/No" questions and checklists were developed. Combinations of these different measure types proved particularly useful. Thus, a Likert scale question followed by a checklist allows the instructor to assess a semi-general question and provide more detail on aspects of performance that were not adequate by selecting elements from a list. These measures are intended to be rated in real-time while an exercise is underway. Therefore, it was particularly important to develop measures that were concise and used operational language as much as possible in both the questions and the behavioral anchors. After analyzing all of the interview data, we had developed 161 performance measures for review by the SMEs.

During the third knowledge elicitation session, we reviewed and revised all of the performance measures with the SMEs according to the following criteria:

- **Relevance:** Does the measure assess an aspect of performance that is important for mission readiness?
- **Observability:** Does the measure assess a behavior that is truly observable?
- **Question wording:** Does the measure make sense to other SMEs?
- **Scale type:** Is the scale used appropriate for differentiating behavior?
- **Scale wording:** Do the behavioral anchors make sense to other SMEs?

When appropriate and based on input from the SMEs, modifications were made to the measures, resulting in a final list of 132 observable performance measures for assessing the performance of the leader during motorized patrol operations. An example of a PM and associated mission phase and PI can be seen in Table 2.

Table 2. Example MPO Mission Phase, PI, and PM

Mission Phase	Conduct: React to Contact
Performance Indicator	Properly communicate contact report
Performance Measure	Does the patrol leader properly communicate a contact report? 1: Talks too much to higher, doesn't lead unit or fails to communicate with higher 3: Attempts to communicate with higher and vehicle CMDRs simultaneously 5: Quickly communicates essential information to higher and then assumes command of unit

Implementation of the Performance Measures in a Rating Tool

In order to be truly usable by instructors, the performance measures developed using the COMPASS Methodology had to be implemented in an easy-to-use rating tool. We chose to modify an existing assessment tool, SPOTLITE, to fit the needs of the SOI-E instructors (MacMillan et al., In press).

While 132 performance measures were developed for MPO, we do not expect that all 132 measures will be rated in each training exercise. Not all of the training exercises, for example, will require the MPO leader to evacuate wounded personnel, one of the mission phases. Additionally, the simulation software used

allows instructors to add events to the simulation in real time. For example, instructors can respond to the trainees' behavior by inserting additional, unplanned, improvised explosive devices (IEDs) once the simulation exercise has begun. It was important, therefore, that the rating tool provide some flexibility to the instructors with regards to selecting the performance measures that should be rated. We used an iterative process of design and development to expand SPOTLITE to allow the instructors to add sets of performance measures according to the different events the trainees may be required to respond to. This allows SPOTLITE for Marine Motorized Patrol (MMP) to be useful across scenarios and training environments. To support ease-of-use and portability, SPOTLITE MMP was installed on a Panasonic Toughbook Tablet PC (see Figure 1).

Figure 1. SPOTLITE MMP displayed on a Panasonic Toughbook PC.

EVALUATION DURING FIELD AND SIMULATION-BASED EXERCISES

After developing the performance measures and implementing them in SPOTLITE MMP, we had the opportunity to evaluate their use during both simulated and field motorized patrol exercises held at SOI-E as a part of their Infantry Squad Leaders Course.

Evaluation Goals

Our primary goal for this evaluation was to assess the measures and their implementation in SPOTLITE MMP. To that end, we generated three hypotheses:

- **Utility in performance measurement.** The ability of SPOTLITE MMP measures to assist instructors in evaluating the performance of patrol leader trainees.
    - Hypothesis: Structured performance measures will assist instructors to assess trainees' performance during simulated and live MPO training, as measured by survey responses by the instructors.
- **Utility during debriefs.** The ability of the SPOTLITE MMP performance data to assist instructors in providing feedback to patrol leader trainees during debriefs.
    - Hypothesis: Debriefs given by the instructors using the SPOTLITE MMP performance data will be perceived as useful by both trainees and instructors, as measured by survey responses by the instructors and trainees.
- **Usability of the assessment tool.** The usability of SPOTLITE MMP.
    - Hypothesis: SPOTLITE MMP would be easy to use, as measured by survey responses by the instructors.

Method

Demographics

Four instructors participated in the evaluation, which included both a period of SPOTLITE MMP use and a post-exercise survey. One instructor used the tool but was unable to complete the survey and his data will not be included in subsequent analyses. The three remaining instructors were Staff Sergeants. The average age was 33.5 (max = 38, min = 29), the average years of service was 14.3 years (max = 17, min = 11), and the average number of combat deployments was 1.7 (max = 2, min = 1).

Forty trainees were enrolled in the Infantry Squad Leaders Course. The average trainee age was 23 (max = 33, min = 19), the average years of service was 3.6 (max = 10, min = 2), and the average number of combat deployments was 1 (max = 3, min = 0). Ten of the trainees had not been on any combat deployments.

Evaluation Design

The trainees were separated into two groups, one that received simulation-based training in the SOI-E Advanced Infantry Training Company (AITC) Deployable Virtual Training Environment (DVTE) Lab prior to the field exercise and one that did not. Both groups received three days field training, composed of one motorized patrol each day. Logistical considerations ensured that each group received similar, but slightly different, training on each day. The field training between days was different. The instructors rated the MPO leader of both groups using SPOTLITE MMP during the one day of simulation-

based training and two of the field exercise training days.

Both instructors and trainees completed surveys related to the use of SPOTLITE MMP for assessing and providing feedback about performance. Each MPO leader trainee answered the survey after the exercise in which they were the leader. There were two MPO leaders for each day of training, each leading 19 other Marines. In their survey, the MPO leader trainees were asked to rate the extent to which they agreed with the following two statements:

1. The AAR provided me with a lot of information about my performance as a patrol leader.
2. I have a better understanding of how to do better next time, based on the AAR.

Both statements addressed the hypothesis concerning the utility of using SPOTLITE MMP performance data during debriefs.

The instructors answered their survey once, after having used SPOTLITE MMP in both the simulated and live training environments. As with the trainee survey, they were asked the extent to which they agreed with the following statements:

Hypothesis addressed: Utility in performance measurement

1. SPOTLITE improved my ability to assess the trainees over previous methods of assessment (i.e., paper and pen).
2. SPOTLITE contributed to my ability to assess the patrol leader's performance during motorized patrol operations.

Hypothesis addressed: Utility during debriefs

3. SPOTLITE enhanced my after action review (AAR), in comparison to a AAR without SPOTLITE.
4. SPOTLITE enabled me to have a more in-depth understanding of the patrol leader's performance.
5. SPOTLITE enabled me to provide better guidance to trainees to improve their performance.

Hypothesis address: Usability of the assessment tool

6. SPOTLITE was easy to use.

All statements on both surveys were rated on a five-point scale, with 'Strongly Disagree' as a '1' and 'Strongly Agree' as a '5.' Respondents were encouraged to make their rating anywhere on the scale between '1' and '5.' The instructor surveys provided room for open-ended comments in addition to the scales.

Analysis and Discussion

Critically, as this was the first time SPOTLITE MMP was used at SOI-E, we were careful to scope our evaluation efforts according to available time and resources at SOI-E. Thus, only descriptive statistics are presented here. They are meant to illustrate *potential trends* (rather than conclusions) and serve as a mechanism for encouraging further evaluation and analysis, particularly with regard to the use of structured performance measurement technologies in both simulated and field training.

Utility in Performance Measurement

Survey questions 1 and 2 were used to evaluate the ability of the SPOTLITE MMP performance measures to assist instructors in evaluating the performance of patrol leader trainees. As seen in Figure 2, the instructors "agreed" or "strongly agreed" with both statements. Summary data for the first and second statements can be seen in Table 3. Open-ended comments included: "To the point" and "Helps keep thought process flowing efficiently."

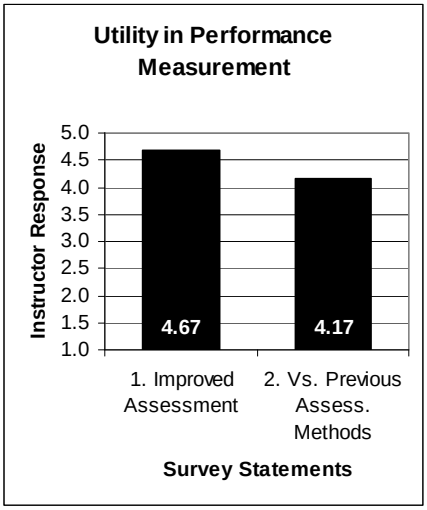

Figure 2. Instructors agreed that the SPOTLITE MMP PMs were useful in helping them assess trainee performance.

The instructors agreed or strongly agreed that the performance measures within SPOTLITE MMP helped them assess the patrol leader's performance. These results indicate, therefore, that the MPO performance measures at least have received support for face validity. Given the congruency of the SPOTLITE MMP measures with observable objectives, it would be expected that the measures would also be valid when compared to an overall criterion variables (in a

controlled setting, across multiple trials), similar to previous results we have received elsewhere (MacMillan et al., In press). This is an important part of truly validating the SPOTLITE MMP measures, but one that was not feasible to assess during this pilot study.

Table 3. Summary data for instructor statements 1 and 2.

Statement	Average	StDev	N
1. Improved Assessment	4.67	0.29	3
2. Vs. Previous Assessment Methods	4.17	0.76	3

Utility During Debriefs

Survey questions 3, 4, and 5 were used to assess the ability of the SPOTLITE MMP performance data to assist instructors in providing feedback to the patrol leader trainees during debriefs. As seen in Figure 3, instructors somewhat agreed that the performance data assisted their ability to conduct debriefs. They strongly agreed that it provided them a more in-depth understanding of the trainee's performance and they agreed that the data assisted them in providing the trainees better guidance. The summary data for instructor statements 3 – 5 can be seen in Table 4.

. Comments from the instructors included: "I was able to rapidly see the area[s] of deficiencies that I could concentrate my attention on for corrective action."

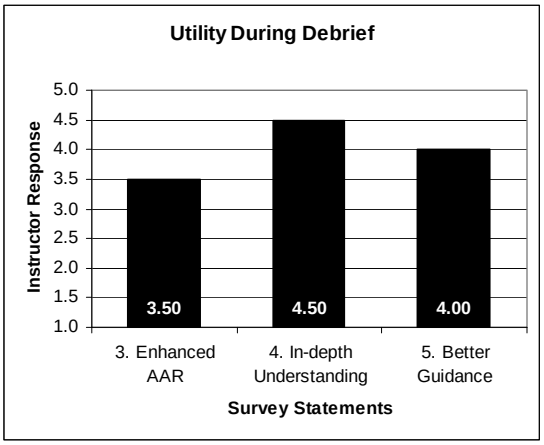

Figure 3. Instructors somewhat agree to strongly agree that SPOTLITE MMP performance data enhanced their debriefs.

Table 4. Summary data for instructor statements 3 – 5.

Statement	Average	StDev	N
3. Enhanced AAR	3.5	0	2
4. In-depth Understanding	4.5	0	2
5. Better Guidance	4.0	0	2

The trainee statements were analyzed according to each training period (one day in the simulation lab and one day each in the field). Because each day of the field training was different, we chose to analyze the survey results separately for those two days, as well. As seen in Figure 4, the trainees mostly agreed that the feedback they received during the debrief in which the instructor used the SPOTLITE MMP performance data provided them with information about their performance. They somewhat agreed that the debrief provided them with a better understanding of how to improve their performance. Table 5 shows the summary data for trainee statements 1 and 2.

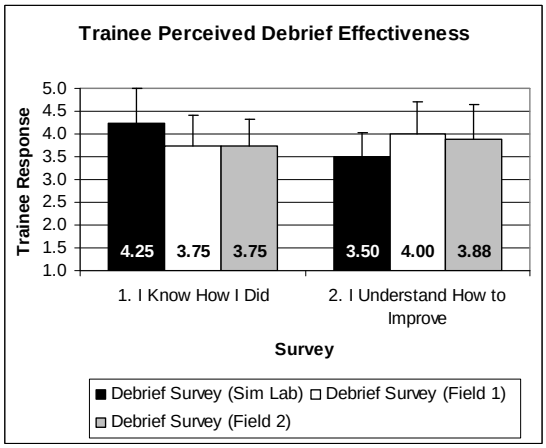

Figure 4. Trainees agreed that debriefs given using the SPOTLITE MMP performance data provided them with more information about their performance.

Table 5. Summary data for trainee statements 1 and 2.

	Sim Lab			Field Exercise 1			Field Exercise 2		
Statement	**Average**	**StDev**	**N**	**Average**	**StDev**	**N**	**Average**	**StDev**	**N**
1. I Know How I Did	4.25	0.35	2	3.75	1.06	2	3.75	0.5	4
2. I Understand How to Improve	3.5	0.71	2	4.0	1.41	2	3.88	0.63	4

Usability of the Assessment Tool

Statement 6, "SPOTLITE was easy to use," assessed the usability of SPOTLITE MMP. The instructors average rating was 4.67 (StDev of 0.29, N = 3). Comments received from the instructors included: "Straightforward," "Dummy proof," and "Easy to comprehend." Based on these results, it appears that we achieved our goal in making an assessment tool that required little upfront training and practice and was, indeed, easy to use. However, conducting a more thorough and directed usability assessment after the instructors have had an opportunity to use SPOTLITE MMP in their everyday activities may yield more insightful and comprehensive usability results.

FUTURE DIRECTIONS

Structured and systematic performance measure collection can provide insight into a variety of aspects of training. It can, naturally, provide instructors the ability to immediately assess the performance of a specific trainee or a team of trainees. Over time, carefully collected performance data can provide insight into the effectiveness of specific training scenarios or curriculum. Additionally, when used in both simulated and live training exercises, this performance data can provide guidance on how to best combine these two training environments, utilizing each to its maximum potential. Below, we described an additional effort we see as being important in the continued development of MPO training and assessment techniques.

Experimentally Validate MPO Performance Measures

As discussed previously, experimental validation of the MPO performance measures developed in this effort is necessary to ensure that their subsequent use in assessing trainee performance is reliable, valid, and sensitive. While the face validity described above is useful and important, only an experimental validation can provide the necessary data to support such claims.

Conducting such an evaluation for the MPO performance measures would be most readily completed in a simulation environment, where the scenarios rated by instructors can be controlled most rigorously. In fact, many simulation environments allow the exercise to be recorded for playback later on, thus providing the exact same scenario to be rated by multiple instructors. In this validation experiment, a variety of recorded scenario exercises would be selected. They should represent all of the mission phases and sub-phases and demonstrate a range of performance across these items, to include highly experienced and less experienced MPO leaders.

In a between subjects design, half of the instructors would rate the performance of the MPO leaders (i.e., all of the selected scenarios) using SPOTLITE MMP and half would use an overall criterion measure only. This would allow us to assess the overall validity of the performance data capture by SPOTLITE MMP. Inter-rater reliability would be investigated by computing the coefficient alpha for each scenario (Nunnally, 1978). This can be investigated overall and by each mission phase and/or sub-phase, as well. Finally, sensitivity would be assessed by investigating the range of the SPOTLITE MMP and overall criterion performance rating averages and standard deviations, with the expectation that highly experienced MPO leaders would receive high ratings and less experienced MPO leaders would receive lower ratings.

CONCLUSIONS

During this development effort and pilot study, we developed a systematic set of performance measures for assessing the performance of patrol leaders in MPO training exercises. Trends in this initial data indicate that the measures were deemed useful by both instructors and trainees, who perceive the measures as adding value to both simulation-based and field training. Additionally, we developed an assessment tool that allows instructors to take ratings during the course of an MPO training exercise. Results indicate that the assessment tool, SPOTLITE MMP, is easy to use. Future research with the performance measures should include a formal evaluation of the performance measures with respect to inter-rater reliability, validity, and sensitivity.

ACKNOWLEDGEMENTS

This work was performed under contract with PMTRASYS, with guidance from Mr. Martin Bushika and Mr. John Styer. We would like to thank the instructors at SOI-E AITC for their support and participation in all of the activities described in this report, with particular thanks to the Infantry Squad Leaders Course instructors. Thanks to Dr. Shawn Weil and Dr. Cullen Jackson for their thoughtful comments during the development of this paper. We would also like to thank Dr. Winston Bennett for supporting the initial development of COMPASS and SPOTLITE, which provided a strong foundation on which to conduct the work with PMTRASYS and SOI-E.

REFERENCES

Biddle, E., Keller, M., Pitz, R., & Nixon, A. (2005). *Implementing an Automated Performance Assessment Capability for the Joint Training Special Event (JVTSE).* Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2005, Orlando, FL.

Dieckhaus, D., 1stLt. (2006, October 2006). Reevaluating Convoys and the IED. *Marine Corps Gazette,* 15 - 18.

Gagne, R. M., Briggs, L. J., & Wager, W. W. (1992). *Principles of Instructional Design* (Fourth Edition) Orlando, FL: Harcourt Brace Jovanovich College Publishers.

Lyssy, B., & Munden, W. (2006). *The Marine Corps Deployable Virtual Training Environment Virtual Fire Support Trainer (DVTE-VFST).* Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2006.

MacMillan, J., Entin, E. B., Morley, R. M., & Bennett Jr., W. R. J. (In press). Measuring Team Performance in Complex and Dynamic Military Environments: The SPOTLITE Method. *Military Psychology*.

Nunnally, J. (1978). *Psychometric theory*. New York: McGraw Hill.

O'Bea, M., Bell, C., & Crabtree, J. (2006). *Spiral Development of Virtual Solutions for Convoy Defense Training.* Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2006, Orlando, FL.

Phillips, J. K., Shafer, J., Ross, K. G., Cox, D. A., & Shadrick, S. (2005). *A Behaviorally Anchored Assessment Tool to Measure Tactical Thinking Proficiency.* Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2005, Orlando, FL.

Roby, G. H., Jr., & Pedersen, C. J. (2006). *Combined Arms Planning Tool (CAPT) Training Fire Planning*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2006.