SANDIA REPORT
SAND2012-8812P
Unlimited Release
Printed September 2012

Enhanced Training for Cyber Situational Awareness in Red versus Blue Team Exercises

Armida Carbajal, Susan Stevens-Adams, Austin Silva, Kevin Nauer, Benjamin Anderson & Chris Forsythe

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND 2012-8812 P
Unlimited Release
September 2012

Enhanced Training for Cyber Situational Awareness in Red versus Blue Team Exercises

Armida Carbajal
Human Factors and Statistics

Susan Stevens-Adams
Risk and Reliability Analysis

Austin Silva
Cognitive Modeling

Kevin Nauer
Cyber Security Technologies

Benjamin Anderson
Assurance Technologies and Assessment

Chris Forsythe
Cognitive Modeling

Sandia National Laboratories
P.O. Box 5800
Albuquerque, New Mexico 87185-1188

Abstract

This report summarizes research conducted through the Sandia National Laboratories Enhanced Training for Cyber Situational Awareness in Red Versus Blue Team Exercises Laboratory Directed Research and Development project. The objective of this project was to advance scientific understanding concerning how to best structure training for cyber defenders. Two modes of training were considered. The baseline training condition (Tool-Based training) was based on current practices where classroom instruction focuses on the functions of a software tool with various exercises in which students apply those functions. In the second training condition (Narrative-Based training), classroom instruction addressed software functions, but in

the context of adversary tactics and techniques. It was hypothesized that students receiving narrative-based training would gain a deeper conceptual understanding of the software tools and this would be reflected in better performance within a red versus blue team exercise.

A study was conducted in which participants (students) received either tool-based or narrative-based training. Following classroom training, three teams were formed to compete in an exercise requiring they use the selected software tools to solve various challenges associated with threats to network security. For this study, training focused on a set of software tools selected because they represented commonly used tools within operational cyber settings. The three teams were: (1) Tool-Based, for which each participant received tool-based training; (2) Narrative-Based, for which each participant received narrative-based training and (3) Combined, for which two participants received tool-based training and two received narrative-based training.

Statistical comparison of team performance did not reveal significant differences between the three teams, although this was not unexpected given the small number of participants. However, when each team was asked to provide their collective account of the multi-layered attack confronted during the exercise, there were clear differences between the teams. In particular, the Narrative-Based team recognized the spatial-temporal relationship between events and constructed a timeline that was a reasonable approximation of ground truth. In contrast, the Combined team produced a linear sequence of events that did not encompass the relationships between different adversaries. Finally, the Tool-Based team demonstrated little appreciation of either the spatial or temporal relationships between events. These findings suggest that students receiving narrative-based training were able to use the software tools in a way that allowed them to gain a much greater level of situation awareness.

Secondary measures revealed a statistically significant relationship between the personality attribute Emotional Stability and performance in the cyber exercise. Furthermore, the cognitive aptitude measure Comprehension Span, which assesses an individual's verbal comprehension skills, showed a statistically significant correlation with cyber exercise performance.

While the study failed to clearly demonstrate the expected benefits of Narrative-Based training approaches in training for the use of cyber software tools, benefits were observed with respect to improving the ability of students to interpret and comprehend a multi-layered cyber attack. Additionally, insights were gained concerning individual attributes that contribute to the success of cyber defenders.

ACKNOWLEDGMENTS

The authors would like to acknowledge the technical assistance to this project provided by John Jarocki, Theodore Reed, Jonathan T. McClain, Jonathan Mandeville, Kasimir Gabert and Alex Quintana. This project would not have been possible without their contributions.

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

CONTENTS

List of Figures

List of Tables

NOMENCLATURE

BFI	Big Five Inventory
CTF	Capture the Flag
IR	Infrared
MS	Microsoft
Tracer FIRE	Tracer Forensic Incident Response Exercise

1. INTRODUCTION

Situation awareness is essential to effective cyber security analysis and incident response team performance. However, cyber situation awareness is poorly understood. This project sought to help clarify the cyber situation awareness problem, while providing insights that will improve training effectiveness for cyber defenders.

An explosion of new vendor and open source tools has occurred in the past few years to address the growing cyber problem, with U.S. Government enterprise networks and their incident response teams being a primary market. However, these new tools have not always improved the situation awareness of cyber security analysts. Consequently the return on investment has been questionable given the costs of purchase, development and integration of the new technologies.

Nonetheless, cyber security analysts need tools to assist them in fathoming the vast quantities of data and deciphering ever-more sophisticated network attacks. There is need for research to understand why tools that ought to increase the productivity of cyber security analysts often fail to realize this objective. We believe that this failure may be partially attributable to insufficient training, and particularly, the fact that intended users often lack fundamental knowledge essential to effectively use the tools being provided to them. Today, there is no scientific basis for asserting that one mode of training cyber defenders to use software tools is superior to any other mode of training. Likewise, there has been no openly published empirical assessment of students receiving alternative modes of training. The objective of this project was not to compare alternative software tools and no data was collected that reflected on the relative performance or utility of alternative software tools. Instead, through laboratory research employing human performance measurement, the current project scientifically addressed the question of what type of training is needed to maximize the effectiveness of new tools being introduced to improve the situation awareness of cyber security analysts.

The current project employed a suite of network analysis tools comparable to those commonly used in operational cyber settings. This suite of software tools included: Encase Enterprise, Wireshark, IDA Pro, Volatility, Hex Workshop and PDF Dissector. Teams were additionally provided IRC chat as a means for intra-team communications and Plotweaver as an aide in creating a record of events. Two modes of training were considered. The baseline training condition (Tool-Based training) was based on current practices where classroom instruction focused on reviewing the software functionality with various exercises in which students apply those functions. In the second training condition (Narrative-Based training), classroom instruction addressed software functions, but in the context of adversary tactics and techniques. Upon completion of training, participants were evaluated during a Tracer FIRE (Forensic and Incident Response Exercise) simulated blue team exercise. It was hypothesized that students receiving narrative-based training would gain a deeper conceptual understanding of the software tools and that this would be reflected in better performance during the Tracer Fire exercise.

Three hypotheses were tested.

Hypothesis 1: The narrative-based training is different from the tool-based training and will result in better performance in an assessment of students' abilities to use software tools to interpret events associated with a cyber attack.

.

Hypothesis 2: Personality has an effect on team success and dynamics. Certain personality attributes will result in lower team scores.

Hypothesis 3: Cognitive aptitude has an effect on team success. Certain cognitive aptitudes will result in superior team scores.

While research of this nature is commonplace in other high consequence domains (e.g., military operations), there exists little precedent within the cyber security domain. Accordingly, the cyber domain introduces unique challenges. For instance, scenarios must be presented that are unique and somewhat realistic, yet offer equivalent outcome measures of performance. Process measures must be identified and implemented that allow data to be collected in a non-obtrusive manner such that measurement does not interfere with participants exercising the skills and knowledge being measured. Furthermore, outcome and process measures must be identified that are generalizable to and predictable of performance within operational settings. By beginning to address these issues, the proposed project advances the domain of cyber science through development of unique experimental methodologies, while providing a deeper understanding of situation awareness within the cyber domain.

Furthermore, the current study offered an opportunity to collect data regarding secondary research questions concerning the effectiveness of cyber operations. Cyber security is a major challenge for DOE and other government agencies and there has been little scientific study of the human dimension of cyber operations. Through the current study, data was collected that addressed group processes, the relationship between certain cognitive and personality attributes and the behavior and performance of cyber defenders, and the use of narrative in constructing stories to understand, explain and remember events in the cyber domain.

2. METHODS

The study involved two phases. Once subjects had been selected to participate, they were assigned to one of two training groups (Tool-Based or Narrative-Based) for the first phase of the study. Then, for the second phase of the study, participants were assigned to one of three teams for the performance assessment (i.e. Tracer FIRE exercise). Performance comparisons occurred at the level of individual performance and at the level of team performance.

2.1 Subjects

Subjects were recruited from the Sandia National Laboratories workforce. There were a total of 13 subjects who met the criteria for participation. Subjects were required to: (1) be 18 years or older, (2) have a background in computer science, (3) have an interest in cyber security/cyber incident response; (4) have not participated in any prior Tracer FIRE events and (5) be available on the designated dates for five full days of training and three full days to participate in the Tracer FIRE evaluation exercise.

Individuals who met the requirements for participation were asked to complete a consent form and fill out a pre-screening questionnaire to ensure that they possessed adequate knowledge to participate. The consent form was returned when subjects arrived to begin the first day of the experiment and any questions subjects had concerning their participation were answered at that time.

2.2 Procedure

Subjects were first asked to complete two questionnaires. One was a demographic questionnaire which asked the participant their age, gender, educational background, years of experience with computers and years of experience with cyber security (if any). Next, participants were asked to complete a detailed questionnaire assessing general computer security and cyber incident response skills. This information was later used to assign individuals to the two training conditions and subsequently, place individuals into teams for the Tracer FIRE exercise. The objective was to assure the three teams competing in the Tracer FIRE exercise were relatively balanced with respect to the knowledge and experience of team members.

2.2.1 Training

Subjects were assigned to either the Tool-Based or the Narrative-Based training conditions, with seven subjects assigned the Tool-Based and six subjects assigned to the Narrative-Based condition.

Subjects assigned to the Tool-Based training condition received 3 days of training focused on the functions incorporated into the tools and the mechanics of using the tools. This training involved relatively little information concerning adversary tactics and techniques. This training was comparable to training commonly provided by software vendors and included canned examples showing how the tools work, with relatively little emphasis upon the application of tools to real-world problems.

Subjects assigned to the Narrative-Based training condition received 3 days of training emphasizing the theory of adversary tactics, application of the tools, and a detailed understanding of the role as a cyber incident responder. This training involved little consideration of the functionality of tools used for conducting network analysis. The training was structured in a manner that sought to help students comprehend the complex ideas and information in a form that was personal and formed relationships between their prior knowledge and personal experiences.

The two training groups were combined for 2 additional days of training which addressed details concerning the use of the selected software tools. This training was not as extensive as that provided in the tool-based training and emphasized the knowledge participants would need to solve the challenges in the Tracer FIRE exercise. This training also included information about what would be expected of them during the Tracer FIRE exercise (i.e., how to work as a team, how the challenges would be structured, etc.).

In summary, all of the participants received the same amount of training (40 hours), although one group received intensive Tool Based training (i.e., Tool-Based training condition) followed by a

general basic tool use course while the second group (Narrative-Based training condition) received applied training followed by the same general basic tool use course.

2.2.2 Tracer FIRE Exercise

The Tracer FIRE exercise served as the testing session during which the effectiveness of the alternative modes of training was assessed.

Following the 5 days of training, the participants were placed in one of three teams for the Tracer FIRE exercise. The teams were (1) Tool-Based, composed of four subjects that had been in the tool-based training condition, (2) Narrative-Based, composed of four subjects that had been in the Narrative-based training condition and (3) Combined, composed of 3 subjects that had been in the Tool-Based training condition and two subjects that had been in the Narrative-Based training condition.

Each team was asked to solve multiple challenges to receive points with the scores for each team continuously displayed and teams encouraged to compete with one another. The challenges were built around a coordinated series of events involving the same multi-level attack upon the host network of each team. Challenges required teams to use the software tools addressed during training to analyze network traffic. This provided the basis for their interpreting events and establishing overall situation awareness. Points were awarded on the basis of successfully answering challenge questions concerning specific aspects of the attack, as well as their ability to form an accurate picture of the overall pattern of events (i.e., situation awareness).

2.2.3 Secondary Measures

During the Tracer FIRE exercise, the participants were asked to wear sociometric badges which recorded their verbal communication and spatial-temporal position. The objective was to collect data concerning the frequency, duration and patterns of verbal communication that occurred within the teams.

Subjects were asked to complete a personality assessment that consisted of the Big Five Inventory (BFI) from the website www.similarminds.com. Subjects were also asked to perform three cognitive tasks. The cognitive tasks have been used in previous studies and address different cognitive aptitudes associated with adaptive thinking and decision making. The objective was to assess whether these same aptitudes correlated with performance for the cyber defender tasks. The cognitive tasks are described below:

Syllogism. This task is a measure of reasoning. Participants were given a logical argument in which a proposition is inferred from a set of premises. The participant must decide whether the argument is logical or not. An example of this is: “All rabbits have fur (premise), Some pets are rabbits (premise), therefore, some pets have fur (proposition)”. Participants were asked to indicate whether the proposition was true given the premises.

Comprehension span. This task was a measure of verbal comprehension and associated memory recall. The participant saw sentences and had to indicate (using the mouse buttons) whether the sentence made sense or not. After a series of sentences, the participant was asked to recall (using the keyboard) the last word of every sentence in order.

Mental rotation. This task is a measure of visual-spatial ability and mental flexibility. Participants were presented with a series of 20 pairs of figures. The task was to indicate whether the two figures corresponded to the same object. The number correct that were classified in 60 seconds was taken as a measure of mental rotation ability.

Finally, at the beginning of the Tracer FIRE exercise, subjects were told that there was a story embedded within the upcoming series of challenges. Furthermore, it was their task to discover this story as they solved the various challenges. At this time, it was encouraged that teams pay attention to cues associated with the stories and take notes to help them later piece together these cues. Then, at the end of the Tracer FIRE exercise, teams were given 30 minutes during which their task was to construct an illustration depicting their interpretation of events and the underlying story.

3. RESULTS

3.1 Descriptive Statistics

Subjects were assigned to teams in a manner that provided a relative balance in the skills and experience of the individual team members. With respect to the questionnaire assessing general computer security and cyber incident response skills, the sums of the test scores for each team were Tool-Based training *Team 1 = 354*, Combined Type Training *Team 2 = 374*, and Narrative-Based training *Team 3 = 347*. Table 1 provides the detailed scores.

Demographic data was collected from each participant (see Table 2) and education and average number of years of computer experience was used to assign the participants to teams. Since the individuals did not have equivalent education levels, the fairest combinations were selected for each team. However, in terms of the number of years of computer experience, the Narrative-Based team was intentionally given a disadvantage; this team had the least years of experience. As can be seen in both tables, even though the Combined team had one additional member, their total test scores and demographics are roughly equivalent the other teams.

All of the participants were males. The mean age was *22.15* years. While not used in constructing the teams, the Narrative-Based team reported a greater number of CTFs (Capture the Flag) and years of hacking experience than the other teams.

Table 1. Combined Scores for General Computer Security and Cyber-Incident Response Questionnaire

Training Type	Team	Subjects (n)	Binary Ops Assemby Total	Windows RE Memory	Stack & Heap Exploits Total	Linux	Sum of Grand Total Test Score
Tool-based	1	4	53	24	34	14	354
Combined	2	5	35	13	35	20	374
Narrative	3	4	37	6	40	20	347

Table 2. Summary of Demographic Data of Teams

Training Type	Tool-based	Combined	Narrative	Grand Total
Team	**1**	**2**	**3**	**-**
Subjects (n)	4	5	4	13
Mean Age	22	23.2	21	22.15
s	1.414214	5.3572381	3.162278	3.67
Mean Yrs Computer Experience	11.5	11.8	7.75	10.46
s	5.196152	4.7116876	3.095696	4.48
Mean Yrs Cyber Sec	1.625	2.6	2.875	2.38
s	1.493039	4.2190046	3.520772	3.14
Mean Yrs Hacking	1.25	0.0003333	2.5	1.36
s	1.258306	0.0005774	3.535534	2.3
Mean CTFs attended	1.25	1.25	3	2
s	1.892969	2.5	2.44949	2.25
Highest Level of Education Completed (n)				
H.S.	2	3	3	2.666667
B.S.	2	1		1.5
M.S.		1	1	1

s= standard deviation

3.2 Training Type and Team Differences

As shown in Figure 1, the Narrative-Based team received the most points followed by the Tool-Based and Combined teams, respectively. This was also reflected in the average number of points received by team members; members of the Narrative-Based team individually scored more points on average than members of the other two teams (See Figure 2).

Figure 1. Total Score Obtained by Each Team. Team 1 is the Tool-Based, Team 2 the Combined, and Team 3 the Narrative-Based.

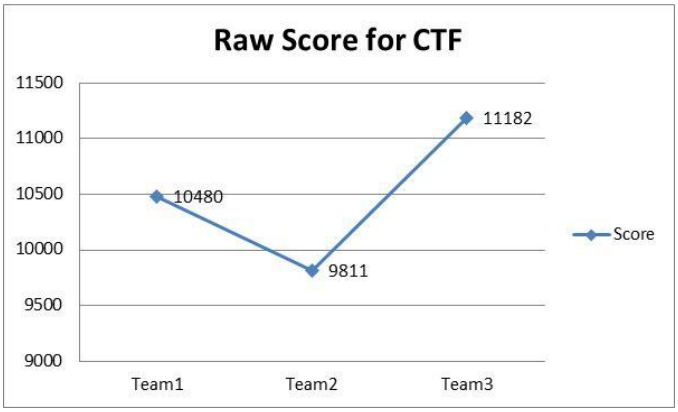
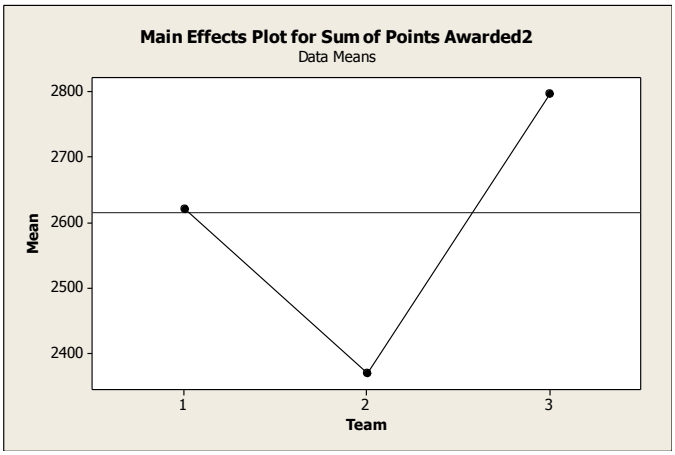


Figure 2. Average Points Earned by Individuals for Each Team.



A general linear model ANOVA with two factors was conducted to determine if there was a "training type" or "team" effect on the number of points obtained by teams. There were two levels in the training type: Narrative-Based or Tool-Based training. There were three levels in the team factor: Team 1 (Tool-Based), Team 2 (Combined), and Team 3 (Narrative-Based) training.

The "training type" factor was not significant, with an $F(1,8) = 0.82$, $p\text{-value} = 0.393$. The "team" factor was not significant, with an $F(2,8) = 0.57$, $p\text{-value} = 0.585$. There was no

statistical difference between team scores (i.e., "success") based on the training type or team. The two training conditions had no effect on overall team success with an $R^2 = 14.04$ and R^2-*adjusted = 0.00.*

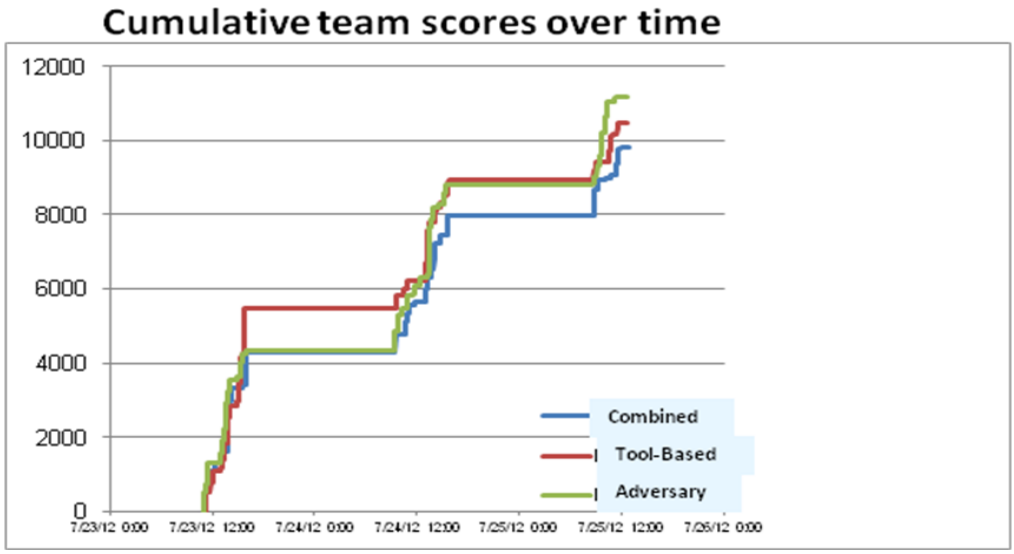
Table 3. ANOVA for Total Points Awarded by Training Type and Team

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Training Type	1	144671	744510	744510	0.82	0.393
Team	2	1046506	1046506	523253	0.57	0.585
Error	8	7291962	7291962	911495		
Total	11	8483139				

S = 954.723 R-Sq = 14.04% R-Sq(adj) = 0.00%

Figure 3 provides the cumulative score for each team across the two days they participated in the exercise. It was noted that initially, the Tool-Based team outperformed the other teams. However, as the exercise progressed and the challenges became more difficult, the Narrative-Based team performed best.

Figure 3. Cummulative Scores of Teams over the Course of the Exercise.



3.3 Team Success and Personality

The BFI data was analyzed to determine if personality measures were associated with team success. One subject withdrew from the study and another subject opted out of the personality assessment portion of the study. Therefore, only 11 participant's data was analyzed. A correlation matrix (Figure 3) was calculated to determine if there was multicollinearity (Penney et al., 2011). The variables showed random scatter and no significant correlation.

A stepwise regression was conducted to determine if any of the variables were significant. *Alpha =0.15,* was set for selection in the stepwise regression. The variables that went into the starting model were: Extroversion (Ext), Orderliness (Ord), Emotional Stability (Emo), Accommodativeness (Accom) and Inquisitiveness (Inq). Based on prior research within this domain, the variables 'TimeTotal' (i.e., total amount of time spent working on challenges) and 'Inq' were also included in calculating the model.

The final model included TimeTotal, Inq, and Emotional Stability. There were no departures from normality or outliers, and the residuals displayed constant error variance, with the error terms normally distributed. These data indicate that participants fell within the range that would be considered normal within the overall population and, therefore, results cannot be attributed to individual subjects with extreme scores on the Inquisitiveness or Emotional Stability personality dimensions.

TimeTotal was not significant, but was included in the model as $\beta_{1.}$ Inquisitiveness and Emotional Stability were significant and the model had an $R^2 = 58.87$ and R^2*-adjusted = 41.25*. The regression coefficients were $\beta_0 = -7491$, $\beta_1 = 1.148e\text{-}12$ with a *t-vaule =1.83, p-value < .1093,* Inquisitiveness was marginally significant at $\beta_2 = 63$ and a *t-value =2.02, p-value=.083* and Emotional Stability $\beta_3 = 88$ and a *t-value =2.98, p-value=.021*.

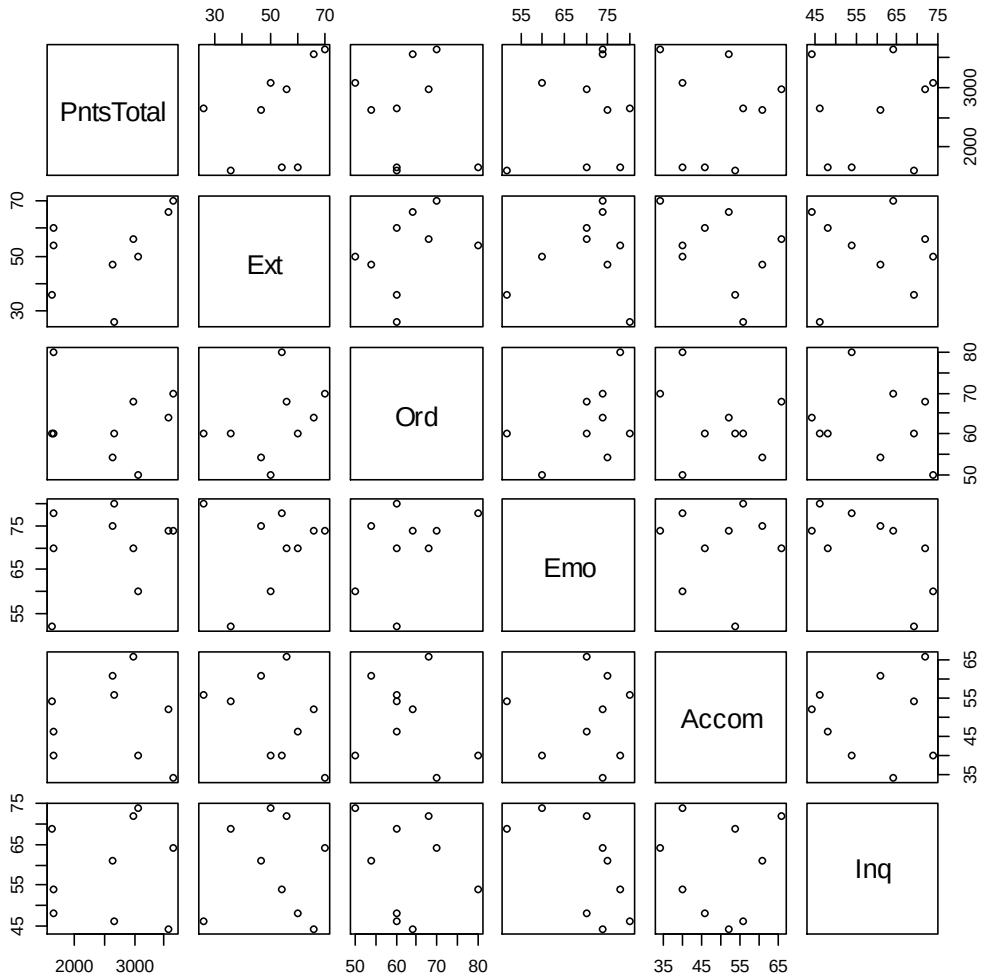
(1) Regression Formula for Team Success: $\hat{y} = -7491 + 1.148e - 12(x_1) + 63(x_2) + 88(x_3)$

Contrary to the '*a-prior'* hypothesis, the most important variables were not total time spent on the challenges (TimeTotal) and Inquisitiveness but instead was Emotional Stability. The stepwise regression was repeated without TimeTotal and Inquisitiveness. Only Emotional Stability was included in the final model.

The regression coefficients were $\beta_0 = -674.5$, $\beta_1 = 47$ with a *t-vaule =1.98, p-value = .078,* Emotional stability was marginally significant and the model had an $R^2 = 30.45$ and R^2*-adjusted = 22.72.*

(2) Regression Formula for Team Success: $\hat{y} = -674.5 + 47(x_1)$

Figure 4. Analysis of Multicolinearity for Personality Measures.



Ext = Extroversion
Ord = Orderliness
Emo = Emotional Stability
Accom = Accommodativeness
Inq = Inquisitiveness (Intellectual Curiosity)

3.4 Team Success and Cognitive Factors

The three cognitive tasks, Mental Rotation (MRScore), Comprehension Span (CompS) and Syllogism (Syllo), were analyzed to determine if they were associated with team success. One subject withdrew from the study and 3 subjects opted out of the cognitive task portion of the study. Therefore, only 9 participant's data was included.

A stepwise regression was conducted to determine if any of the variables were significant. *Alpha =0 .15*, was set for the selection in the stepwise regression. The variables that went into the starting model were: Mental Rotation (MRScore), Comprehension Span (CompS) and syllogism (Syllo).

The final model included CompS. There were no departures from normality, no outliers, the residuals displayed constant error variance, and the error terms were normally distributed. Thus, the results could not be attributed to individual subjects who exhibited extreme scores, as all subjects were within the range that would be considered normal for the population.

CompS, was significant and the model had an $R^2 = 47.25$ and R^2*-adjusted = 39.71*. The regression coefficients were $\beta_0 = 1459$, $\beta_1 = 32$ with a *t-vaule =2.50, p-value <0 .041*.

(3) Regression Formula for Team Success: $\hat{y} = 1459 + 32(x_1)$

3.5 Narrative Construction

When teams were asked to prepare an illustration describing the story underlying the various events encompassed in the TracerFIRE exercise, each team began by using pencil and paper to outline the events. However, each team took a very different approach that is believed to reflect their overall situation awareness. Figure 4 shows ground truth depicted using the Plotweaver tool. As shown in Figure 4, there were multiple actors who intersected one another at key points in time. This was a multi-layered scenario that unfolded over time and it was not expected that any of the teams would be able to fully deduce all the relationships that occurred across time and space.

Figure 5 shows the illustration prepared by the Narrative-Based team. It is apparent that this team failed to deduce many of the relationships between actors and events. However, this team did recognize five separate plot lines that loosely corresponded to those depicted in the ground truth storyline. Likewise, they recognized seven of the thirteen points at which the plotlines intersected one another. These two measures are believed to be indicative of the team's overall situation awareness which, as discussed below, was superior to that of the other two teams.

Figure 5. Plotweaver Depiction of the Ground Truth for the Scenario Presented in the Tracer FIRE Exercise.

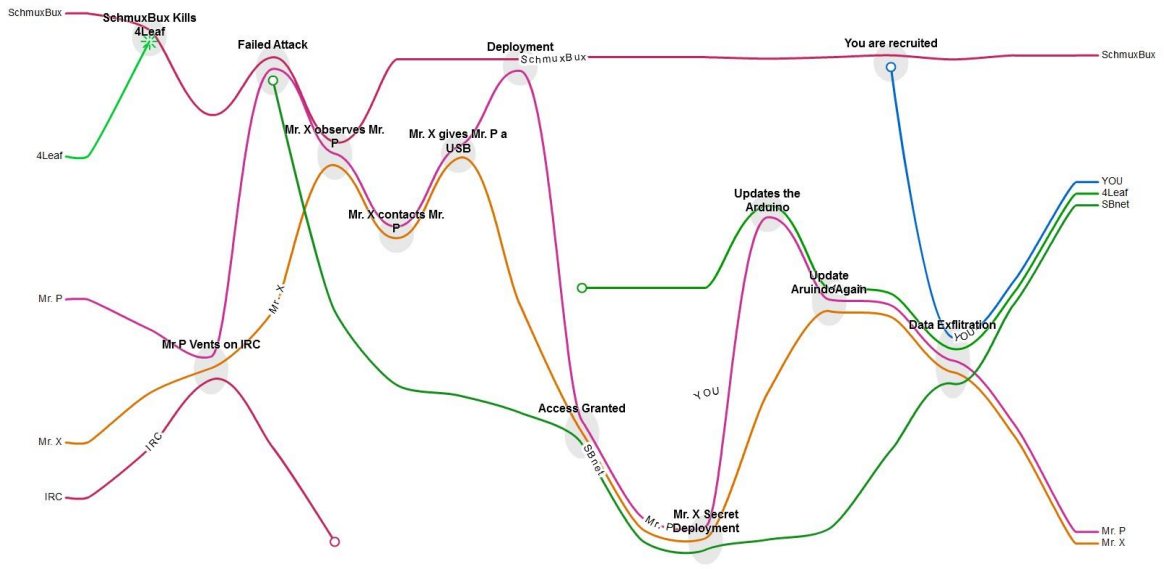
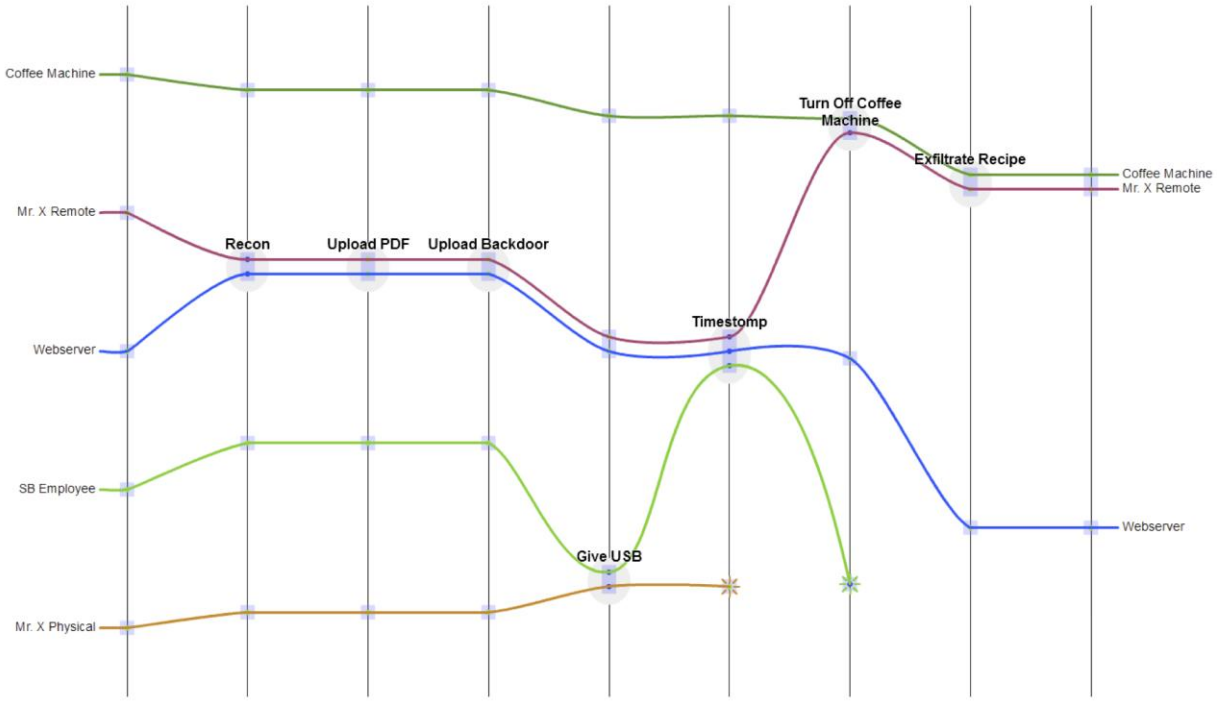


Figure 6. Plotweaver Illustration Prepared by the Narrative-Based Team



The illustration developed by the Combined team is show in Figure 6. This team deduced plotlines that loosely corresponded to four of the five plotlines within the ground truth depiction. Likewise, this team recognized the sequential development of events across time. However, it is striking that this team did not recognize any of the points where the individual plotlines intersected with one another. In fact, in both their hand-drawn illustration and their verbal account, the Combined team presented a linear sequence of events that did not involve any interactions between events, or individual actors. This team pieced together a story involving four separate actors that, for the most part, operated independently, when, in fact, the actors operated in concert with one another and this was a key element to interpreting the overall sequence of events. While this team clearly grasped the temporal structure of events, as well as the importance of individual actors, they were unable to deduce the relationships between different actors that were evidenced through their interactions as the scenario unfolded.

The Tool-Based team produced an even more impoverished illustration than either the Narrative-Based or Combined teams. Their illustration is shown in Figure 7. They recognized three of the five plotlines. Yet, they recognized none of the relationships between the separate plotlines and two of the three plotlines that they did recognize consisted of a single event. Furthermore, their depiction captured none of the relationships between events or the relationships between different actors. Each member of this team seemed to have deduced one or more elements of the story independently; however, as a team, they were unable to put these elements together and did not seem to recognize that there was a coordinated action being taken by the adversaries.

Interestingly, it was noted that all three teams deduced about the same number of story elements. During the Tracer FIRE exercise, there were specific challenges that, if successfully completed, teams learned a key element of the storyline. The points awarded for challenges that provided elements of the storyline are shown in Table 4. While the Narrative-Based team did earn the most points in these challenges, there was not a huge difference between the points earned by the Narrative-Based and the other two teams. This indicates that all three teams had many of the key story elements available to thembut only the Narrative-Based team was able to put those story elements together in a way that corresponded to the actual relationships between events.

Figure 7. Plotweaver Illustration Prepared by the Combined Team

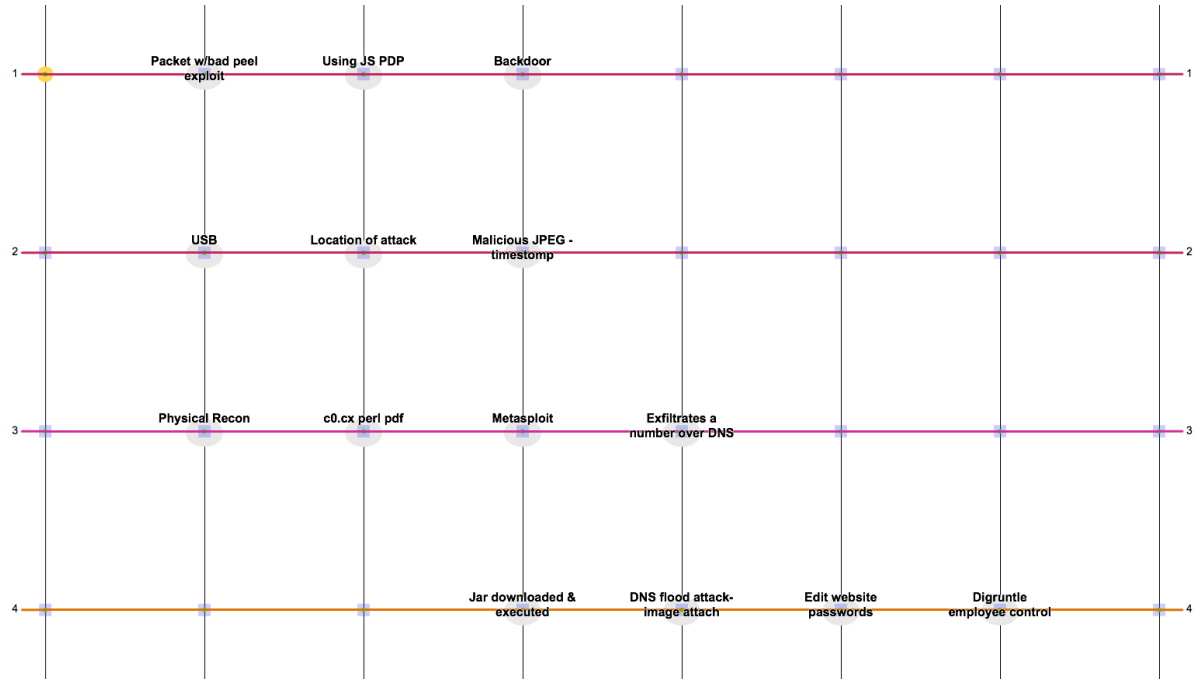


Figure 8. Plotweaver Illustration Prepared by the Tool-Based Team

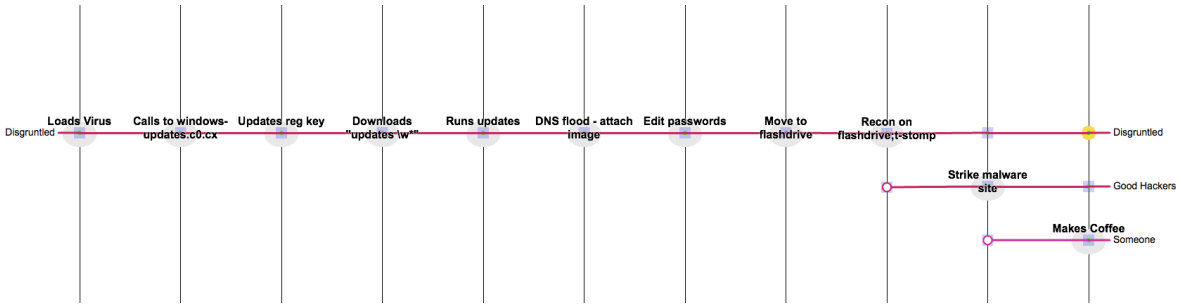


Table 4. Points Awarded for Challenges Providing Elements of the Storyline. Team 1 is the Tool-Based, Team 2 the Combined, and Team 3 the Narrative-Based.

Challenge	Team1	Team2	Team3
4 Leaf 2	200	200	200
4 Leaf 3	0	0	0
4Leaf 1	100	100	100
4Leaf 4	0	0	300
4Leaf 5	0	0	0
Bongo Java Riser 1	100	100	100
Bongo Java Riser 2	200	200	200
Bongo Java Riser 3	300	300	300
Bongo Java Riser 4	0	0	325
Espresso Verdi 1	100	100	100
Espresso Verdi 2	200	200	200
Espresso Verdi 3	300	300	300
Espresso Verdi 4	400	400	400
Iced Americano 1	100	100	100
Iced Americano 2	200	200	200
Iced Americano 3	300	300	300
Iced Americano 4	400	400	400
Mocha-Java 1	200	200	200
Mocha-Java 2	225	0	225
Mocha-Java 3	250	250	250
Mocha-Java 4	275	275	275
Mocha-Java 5	300	300	300
Sumatra 0 'Bonus'	250	250	250
Sumatra 1	300	300	300
Sumatra 2	0	0	0
Sumatra 3	700	700	700
Sumatra 4	0	0	0
Sumatra 5	0	0	0
Sunrise Brew 1	100	100	100
Sunrise Brew 2	200	200	200
Sunrise Brew 3	300	300	300
Sunrise Brew 4	0	325	325
Sunrise Brew 5	0	350	350
Tanzania Peaberry 100	100	0	100
Tanzania Peaberry 200	400	400	400
Tanzania Peaberry 300	300	300	300
Tanzania Peaberry 400	400	400	400
Total	7200	7550	8500

3.6 Team Communications

An initial analysis of data collected from the Sociometric badges was conducted. The objective was to use data regarding interactions between team members as a means to better understand team processes. A substantial quantity of data was obtained, however it was noted that due to the configuration of tables in the laboratory, it is difficult to estimate the reliability of data collection with the Sociometric badges.

The Sociometric badges emit periodic IR pings. Each badge has a receiver that detects IR pings emitted by other badges. Detection of an IR ping is indicative of two individuals being face-to-face. Figure 8 shows the number of face-to-face interactions observed for each pair of team members for each team. It may be observed that teams differed greatly with there being substantially fewer face-to-face interactions for the Narrative-Based team It may also be noted that for both the Tool-Based and Combined training teams, two individuals accounted for most of the face-to-face interactions.

Figure 9. Number of Sociometric badge IR Pings Recorded for Each Pair of Individuals for Each Team

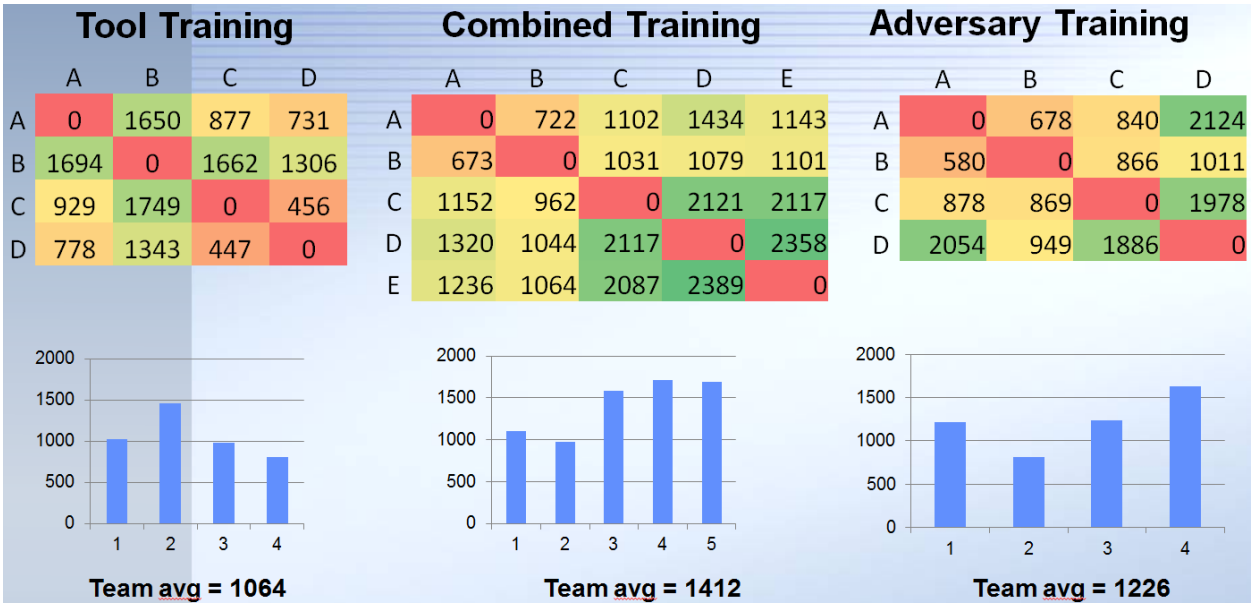
Figure 9 depicts the number of verbal exchanges between each combination of team members for each team. Somewhat different patterns may be observed. For the Tool-Based training team, there appears to be a single individual that dominated discussion. This same individual was one of two that dominated the face-to-face interactions for this team. This suggests that there may have been a single dominant member of the team whose interactions were largely focused on one other member of the team.

With the Combined training team, there were three individuals who seemed to jointly dominate discussion. Interestingly, only one of these two showed a large number of face-to-face interactions. This suggests that much of the interactions occurring with this team did not involve team members making face-to-face contact, while there was one team member who was the

recipient of many exchanges, but initiated few exchanges. In contrast, the Narrative-Based training team exhibited a more equal distribution of communications, with one individual appearing to be slightly dominant and another individual being somewhat less communicative than the other members of the team.

The scope of the current project did not allow for a more thorough analysis of the communication patterns of the three teams. It would have been interesting to relate the communications with both demographics and related data concerning the knowledge and experience of individual participants. Similarly, it would have been worth looking at the moment-to-moment communications in relationship to when challenges were begun and completed, as well as the amount of time working on individual challenges.

Figure 10. Number of Instances of Communications between Team Members for Each Team



4. CONCLUSION

The results from this study provide insights concerning alternative methods for delivering training for cyber defenders, as well as a better understanding of factors contributing to team situation awareness and individual and team performance of cyber defenders. Most notably, this study highlights the importance of the narrative, or the capacity to interpret events and put them into the context of a story, to the effective use of software tools by cyber defenders. Furthermore, the study also illustrates the importance of individual characteristics to the ability of individuals to effectively work together within a cyber incident response team.

With only three teams, it was not possible to demonstrate a statistically significant difference in the performance of the teams receiving alternative modes of training, although the team receiving Narrative-Based training did earn more points than their counterparts. Likewise, on average, the members of the Narrative-Based team individually earned more points than their counterparts on the other teams. While not statistically significant, these results are in the expected direction and are consistent with detailed analysis of overall situation awareness exhibited by the three teams.

It was observed that when asked to describe the overall relationships between events occurring across time and the individuals involved in those events, the Narrative-Based team exhibited a substantially richer interpretation of events that much more closely approximated ground truth. This was in contrast to the other teams who failed to thoroughly grasp the spatial-temporal relationships between actors and events. This is noteworthy because to be effective within an operational context, cyber defenders must go beyond reacting to individual events and must instead correlate events occurring across time and space to deduce the coordinated attacks of adversaries. Thus, in providing students with training that only addresses the functionality of software tools, it is presumed that they possess the knowledge and skills necessary to use the tools in a way that allows them to effectively manage their situation awareness.

Assessments of personality and cognitive factors revealed two variables that were significantly correlated with individual performance during the cyber exercise. With respect to personality, those who exhibited higher scores on the Emotional Stability dimension performed better. Those scoring high on this dimension tend to be more secure and confident, whereas those scoring low exhibit a greater tendency to show unpleasant emotions such as anger, anxiety, depression and vulnerability. It should be noted that while the participants in the current study exhibited a range of scores on this dimension, their scores fell within the range considered normal for the overall population.

There are two important ramifications for the finding that individual performance correlated with Emotional Stability. First, during training, the Emotional Stability of individual students may be expected to affect both the benefit derived from the training experience, as well as the performance during training exercises, such as Tracer FIRE. Thus, it is proposed that mechanisms be employed that allow individual and team performance to be more closely monitored in real-time so that instructors may effectively intervene when students have become non-productive and are struggling. Likewise, in composing teams, it may be beneficial to

combine individuals with varying experience and maturity to provide some degree of scaffolding for weaker team members who may become easily discouraged.

Second, and perhaps more importantly, within operational settings, it may be expected that personnel will exhibit varying levels of Emotional Stability and this will have an indirect, and perhaps direct, effect on their performance. This may be manifested in their capacity to effectively function within teams, as well as their capacity to cope with ongoing stressors. It is uncertain what countermeasures may be most appropriate; however this represents an important consideration given the nature of the Cyber domain where technically qualified personnel are in high demand and many organizations find it difficult to retain their best talent.

A second individual factor that correlated significantly with performance was Comprehension Span. In this task, subjects were presented a series of sentences and after each sentence, they were required to indicate if the sentence made sense. Then, their memory span was tested by requested that they recall the last word in each sentence. To perform well, an individual must have both proficient at interpreting verbal content and possess good short-term memory. Previous studies have shown that individuals who perform well on this measure also perform well in tasks requiring adaptive decision making. Here, adaptive decision making is defined as the capacity to recognize that a strategy is ineffective and thus, there is need to either alter an existing strategy or abandon an existing strategy for an alternative strategy (Abbott et al., 2011). It is proposed that the challenges presented through the Tracer FIRE exercise place similar demands for adaptive decision making upon the participants and that Comprehension Span represents a fundamental cognitive attribute underlying effective performance.

5. REFERENCES

Abbott, R.G. Haass, M., Trumbo, M., Stevens-Adams, S. Hendrickson, S. & Forsythe, C. (2011). Robust Automated Knowledge Capture, SAND 2011-8448, Sandia National Laboratories, October, 2011.

Penney, L.M., David, Witt, L.A. (2011). A review of personality and performance: Identifying boundaries, contingencies, and future research directions. Human Resource Management Review. 21, 297-310.

DISTRIBUTION

1	Benjamin Anderson, 5627		
1	Armida Carbajal, 0431		
1	James (Chris) Forsythe, 1462		
1	John Jarocki, 9516		
1	Jonathan T. McClain, 1463		
1	Kevin Nauer, 9312		
1	Theodore Reed, 5623		
1	Susan Stevens-Adams, 6231		
1	Technical Library, 9536		
1	MS0359	D. Chavez, LDRD Office	1911


Sandia National Laboratories