Evaluating a Multi Agency Cyber Security Training Program Using Pre-post Event Assessment and Longitudinal Analysis

Erik Moore, Steven Fulton, Dan Likarish

Evaluating a Multi Agency Cyber Security Training Program Using Pre-Post event assessment and Longitudinal Analysis

Erik Moore, Steven Fulton, Dan Likarish

Regis University, Center for Information Assurance Studies, Denver, Colorado, USA

{emoore, sfulton, dlikaris} @regis.edu

Abstract. This study presents the context and measured results of cyber security joint cyber defense training exercises. Building on previously published work by the authors that introduced AGILE methodologies, this study analyzes the outcomes of that professional development and laboratory-supported environment. Analysis focuses on the development of specific individual skill levels and generally describes desired multi-agency collaborative capabilities. While all training events do not have sufficient pre-post data to isolate the particular causes of a rise in capabilities, competence in progressively harder levels of capabilities is observed over time in relation to the training components. A comprehensive Personalized Education Learning Environment (PELE) aggregate data of individuals is not presented here, indicators suggest that this would enhance outcomes.

Keywords: cyber·security·defense·training·agile·multi-agency·collaborative·pele·personalized·education·learning·environment·tabletop·simulation·joint·physical·exercise

1. Introduction

The purpose of this study is to determine the efficacy of a joint regional cyber defense training exercise using qualitative assessment and quantitative measures. The impetus for developing a joint regional cyber defense training exercises is an acknowledgement of local entities experiencing cyber attacks such as cyber terrorism, hacktivism, and cyber crime in relation to local, regional, and international events. The frequency, motivation, targets, and types of cyber attacks has expanded significantly [1]. In order to respond to that growing range of cyber-based civil disruption, regional response teams will likely become increasingly necessary in order to provide an agile response to disruptive events, to help under-resourced entities achieve business continuity, and to mitigate damage where possible. This is a type of response is in addition to the array of private services, law enforcement agencies, national defense, and security standards bodies.

Regis University (RU), in partnership with the Colorado Army and Air Force National Guard (CONG) and the State of Colorado (SOC), has completed the first phase of a multi-phase project to build essential and immediate cyber security expertise capacity within the CONG and its information technology (IT) citizen workforce. Their intended cyber-attack response role is similar to their response to a natural disaster like firefighters parachuting out of an airplane, known as "smokejumpers" to put out a forest fire. Successful participants in this training program will be the smokejumpers of the cyber security world. The CONG's IT workforce employs approximately 2000 citizen members with a variety of IT skills and work experience verified through commercial certification. The results of our first year study indicate that the training methodology developed utilizing AGILE methods has improved the ability of the CONG and SOC to respond to attacks against Colorado's critical infrastructure. Based on AGILE development lifecycle principles, the approach we proposed for the Phase I study under a United States National Security Agency/Department of Homeland Security Annex II capacity building grant delivered an executable training partnership model facilitated by collaboration between Academia, the National Guard, and Government. The specific application of AGILE methodology is discussed previously by two of the authors with another collaborator [2]. The study presented here conducted pre and post self-assessment surveys and skill tests reflecting the effect of a multi-year program involving multiple training components, communications with industry, and advanced study of SOC policy.

Because this program relies intentionally on broadly available resources like certification programs, other governmental entities should be able to adapt this model to develop response teams. The training included multiple physical exercises (immersive cyber simulations) and tabletop exercises (collaborative verbal scenario walkthroughs) designed to give real-world substance to more abstract cyber security concepts and integrate physical world consequences to actions performed by the participants.

Items not measured in this research but very important in terms of response capabilities include establishing transitive trust between organizations, a framework for incident response interaction, capabilities to self-organize teams within that framework, and the capacity in the relationship to sustain the empathy and open communications necessary to work through difficult situations. Various activities were designed to provide dynamic group interactions and a common body of knowledge to support discourse and collaborative activities. The exercises incrementally ramped up technical and logistical requirements with what we generally call a crawl-walk-run approach, i.e. starting with easier tasks and working to harder. In addition, the collaborative training space was neither Department of Defense, nor the State of Colorado, releasing constraints for participants to self-organize into a skill-based team structure through rapid discourse rather than extant hierarchies internal to the participating organizations. The significance placed on these capabilities is based on the notion that incident response is the wrong time to introduce responders to each other.exchange

business cards. Below are the activities in our joint training events intended to develop individual skills and team capabilities.

Table 1. Teaching Activities

Teaching Activities	Definition
Simulation	A hands-on immersive scenario which permits instant "reset" of computers, networks, etc. to initial conditions, allows for compression of long term activity into short periods, often provides for lower cost than utilizing real computers, networks, software, protocols, etc, permits ease of scalability, creation of scenarios too risky for "real world" testing. [3]
Physical Exercise Scenario	Activity employing actual technology to engage in a technical challenge in order to develop skill, increase familiarity, and develop team capabilities in relation to a particular active scenario or specific static situation requiring action. This can be "staged" phases with assistance, or timed with measured performance to clarify capabilities. These are immersive experiences that include various levels of social and psychological components to create increasing degrees of validity in relation to the expected real-world events for which one trains.
Tabletop Exercise	Focuses on a primarily verbal walk through team activity of a scenario [4]. Typically, participants role play to review all aspects of a scenario in order to discover challenges to address within the scenario, develop standard operating procedures, identify appropriate team structures, formulate communications plans, and develop default practice. Usually, this does not involve hands-on use of technologies that actual situations would employ.
Lectures	An instructor provides detailed material and students are the receiver of the information. Often, the student is seen as a passive learner, dependent upon the teach to impart what is to be learned. [5]

A relevant approach to cyber security training includes both simulation and physical exercise represent challenge-based learning [6]. This idea of challenge-based learning is a key factor for stakeholders of the training program for validating the capabilities of the participants beyond what is available in certifications, lectures, and other components.

2. Participant Assessment Methodologies

In assessing individual capabilities, pre and post tests were used in the study where data could be gathered. Unlike methods proposed for forming a perfect case, the trainers were obliged to train all participants, and thus a control group could not be established [7]. Therefore the data is analyzed in relation to the longitudinal progress without fully establishing attribution of progress to specific training components over long periods of time. However, general progress of the population can be established by the tests while understanding that there were several factors in the participants'

lives that contributed to their development in capabilities, including requirements for professional certification, continued work activities at their civilian employers, and other National Guard training. The quantitative longitudinal data is therefore represented as general progress of the population and not particularly efficacy of the particular training program. Understanding the professional context of the training clarifies multiple mutually reinforcing technical activities taking place in the lives of the participants.

The analysis presented herein uses what Creswell and Clark refer to as "two phase explanatory design mixed methods research" where quantitative data is used to identify change in a group and qualitative analysis is used to understand that change [8]. The authors conducted pre and post event surveys and interviews characterizing the multi-year regime of physical exercises, lectures and certification exams, all of which advanced the study participants domain knowledge, awareness of SOC policy, and communication level with industry. The joint training leadership included multiple simulation challenges, physical exercise scenarios, and tabletop exercises designed to give real-world substance to more abstract cyber security concepts and integrate physical world consequences to actions performed by the participants.

3. Experiment Structure

To achieve an increase in cyber defense capabilities, the joint training leadership team planned joint events with a set of activities specifically designed to inform participants, exercise their skills, and also facilitate inter-organizational collaboration to validate working capabilities, described in Table 1. This study focuses on measuring of the exercising of skills of the participants.

During information sessions where vendors or organizations presented exercise related materials, participants were encouraged to be interactive in order to move towards functional understanding of the scenario. During physical exercises, participants split up into teams purposefully composed of members from disparate institutions to increase inter-institutional interaction. In addition to addressing the content areas and application usage, the exercise design required participants to practice self-organizing, to close gaps and resolve issues across institutional boundaries in the face of each new challenge. During tabletop exercises, participant teams involved in inter-institutional activity and technical leads walked through scenarios that would require broad multi-institutional cooperation. Specific cases suggesting the need for inter-institutional cooperation included: response to a persistent intrusion at a public utility company, a recurring denial of service attack at a state agency, or a targeted attack against online government services where the local support team is overwhelmed. Each of these scenarios require that inter-institutional teams jointly address the policies of each institution and develop working relationships and methods to achieve functional capabilities while maintaining local responsibilities and adhering to relevant security and operations policies.

During time between joint events, the supervisors of each participating group encouraged participants to work in three areas: 1) Certifications such as CISSP and GIAC, 2) Security applications and relevant procedures, and 3) Rules of engagement based on jurisdictions and inter-institution policies. This study does not definitively separately attribute the sources of skill growth between these directed areas of study and the actual training events. Table 2 shows the categories of content and concepts, along with specific applications and utilities that were identified as warranting usage during physical exercises at the joint events. The research team further divided these categories into specific questions on surveys and tests.

In order to determine the efficacy of the multi-year training program, the authors formulated an assessment strategy, first identifying content and concept areas, and based on that designing a set of pre and post event surveys and tests for understanding in addition to qualitative interviews gathered during events. To provide differentiation between the impact of a specific joint training exercise event and the overall progress of participants in the multi-component program, the authors gathered pre and post assessment of a single joint training event, and then longitudinal assessments at the beginning of every event.

Table 2. Application Categories and Examples

Content area and Concepts	**Outcomes and Assessment**	**Larger Set of Example Applications**
Logging and Monitoring	Composite monitor and visualization to maintain stateful awareness of systems and infrastructure.	Kiwi syslog / Splunk, Alien Vault, Security Onion, SNARE OSSIM
Coding and scripting	Move from user interface based to command line scripting, programming and tool switch orientation	Python, command line interface, powershell
Network, Infrastructure Analysis and System Vulnerability Testing	Cert preparation and testing, skill building, increased training simulator availability for practice.	CISSP (cert), GIA (cert), Wireshark, Kali, Nmap, Metasploit, OpenVAS, Cobalt Strike (Armitage), MobiSEC
Systems audit, compliance and regulation	Audit, compliance	Wapiti, W3A, CISA, CISM (also review of compliance requirements local regulations)
Communication and interpersonal group relationship	Incident handling and involvement of government and commercial	Tabletop exercises, physical exercises, round table discussions with critical

	stakeholders to better define rights and privileges of citizens, National Guard and state officials	infrastructure sectors

The joint training program leadership team members negotiated for their institutions the specific sequence for the delivery of curricular content incrementally, based on an assessment of readiness at each phase of training, and also considered the progression of enabling layers of skill yielded new levels of capability. Once the new curricular content for each event was determined, the particular event was rapidly formulated into a set of challenge exercises covering the categories in the table above. The first event was a network defense challenge similar to the Collegiate Cyber Defense Challenge, with a framework dividing participants into "IT Consultant Teams" required to take over, fix, and defend a mismanaged IT infrastructure as described in an earlier work [2].

The results saw about 10% change from initial population per event ($n > 30$ for the first two events and $n > 13$ where participants were in large part the original population for the 3rd). While this definitely affects that relative percentage of any trend within the core population, "skill gap" closure can still be observed in areas discussed below.

The change in population can be explained by invitation of several groups from outside the core Colorado teams to join in the training exercises, including National Guard representatives from another US state and a Jordanian Military security team. Overall, these representative groups integrated rapidly into the self-organizing teams during physical challenges and expressed value in the presentation of new tools, observation of the tabletop exercises, and participation in discussions to exchange challenges and ideas. Initially, the exercises were composed of Colorado state government, Colorado National Guard, and academic institutions, and quickly attracted interest from IT security industry companies and governmental entities outside of Colorado. This suggested that the model for operating this program might have broad value for other regions. A key point of value in these conversations was that the exercises included multiple simulation, physical exercise scenarios, and tabletop exercises designed to give real-world substance to more abstract cyber security concepts and integrate physical world consequences to actions performed by the participants.

4. Findings

Participants completed evaluation surveys and knowledge tests both at the beginning and at the end of only the first joint event. Longitudinal pre event tests measure variance in capability between the start of each event. One observation was the lack of participants' voluntary participation in the post-event assessment for all events. This could be explained by the fact that participants were tired following the event and did not wish to remain long enough to complete the survey.

The overall trend of data goes upwards over the span of years, while we recorded levels of capability fluctuating up and down within the same event and between events. Interpretation of the pre-post event test data must also take into consideration the compounding nature of the skills introduced over time. This may account for some fluctuation in results. Discrete tracking of particular technical skills suggested that "gap filling" was a stronger aspect of the training events rather than the broad acquisition of skills in new areas.

One possible influencing factor outside of the training events is the study and practice that participants were directed to complete by their respective institutions as part of the overall effort to enhance capabilities. This broader longitudinal acquisition of new skills took place between events, by the efforts of individuals at their jobs, in pursuit of professional certification, and in collaborative study teams. Generally institutional requirements were set with their own program customized for each participant, called here Personalized Experiential Learning Environment (PELE) tracking because it tracks independent professional experience, professional development training, and other related activities. PELE is managed at each these institutions using a spreadsheet for tracking activities against needs. There was no comprehensive system coordinating all related activities across all institutions, except for the annual self-assessment and skill testing data gathered in this study.

The results in Figure 1 suggest that some modest progress was made in most areas with a significant gap being filled in web forensics. The drop in "NMAP - Analysis" score suggests that users became aware of their own skill level as smaller in relation to a larger body of knowledge at this event, likely during the learning challenges.

Tested skills in Figure 2 generally improved over the two-day event, with significant skill gaps improving in relation to a larger body of knowledge. Small changes suggest prior knowledge of the skill at least at the level presented in the challenge event.

Over the span of 2013-2015 as represented in Figure 3, students' perception of skill levels often significantly shifted as a group average. In examples such as "NMAP - Experience" and "Forensic - Storage", relative confidence is often deflated and then rises up over a span the span of years. Other profiles of progress exist within the data, such as areas where there was significant previous experience like "Wireshark" that exhibited consistent skill levels.

While Figure 4 shows some skill gap-filling as identified in areas like backdoors, vulnerabilities, and routers, areas like "ARP poisoning" and "Duplicate DNS entries" suggest that skills emphasized early on may have needed continuing attention to achieve retention. Participants completed the pre event tests using pen and paper individually while proctored in a large classroom.

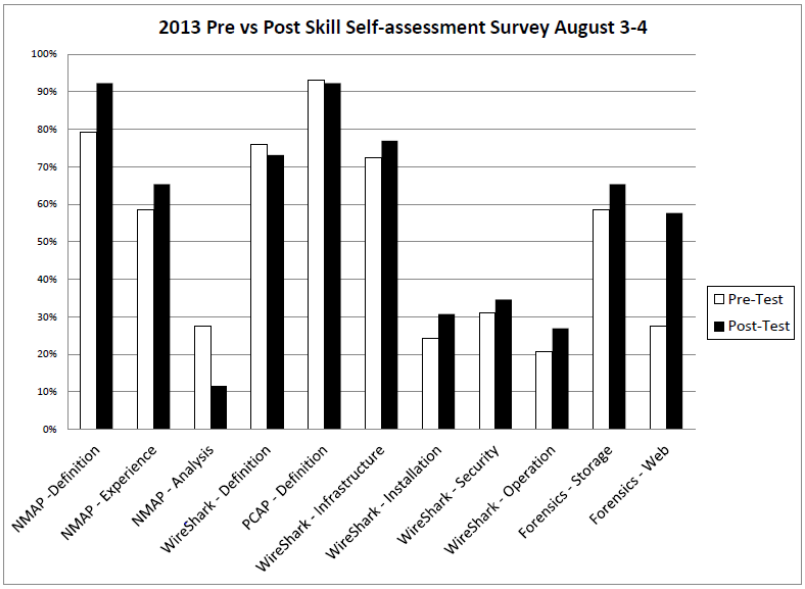

Fig. 1. 2013 User Self Evaluation results plotted an average percentage of self-perceived proficiency across a range of areas (n=30).

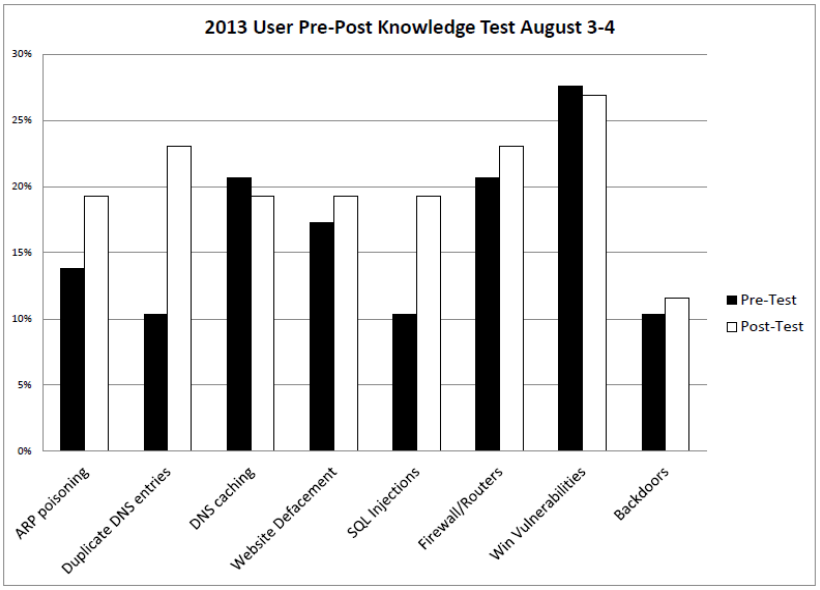

Fig. 2. 2013 Participant knowledge tests prior to and after the challenge event (n=30).

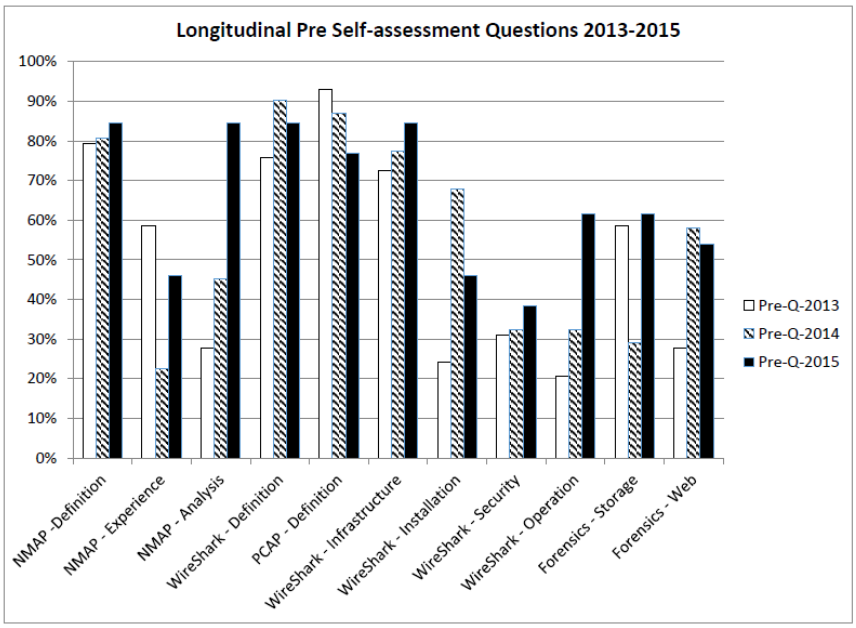

Fig. 3. Longitudinal surveys over three years of self-perceived skills where 2013 (n=30), 2014 (n=30), and 2015 (n=13).

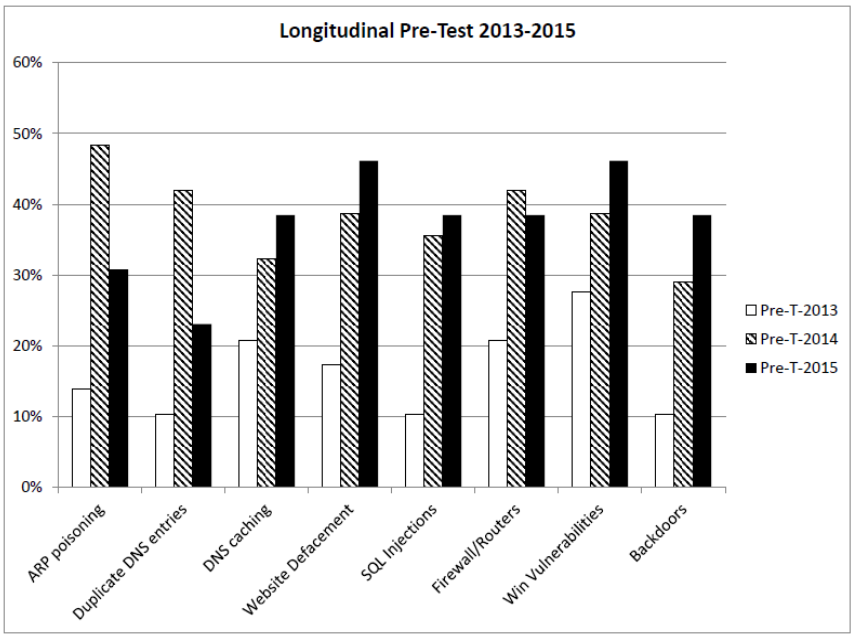

Fig. 4. Longitudinal pre-testing knowledge over the span of three years where 2013 (n=30), 2014 (n=30), and 2015 (n=13).

5. Conclusion

After multiple iterations of surveys and tests we determined through observation of the National Guard working group that an individualized, PELE would benefit the user and manager. The documents that we examined were after action reports, interviews with the managers responsible for the participating institutional team members and a compilation of reports by observers. Not surprisingly the user and manager requirements were similar provide accurate and real-time situational awareness of strengths and weaknesses The managers were primarily interested in how best to respond to an immediate threat given the available workforce and up to date competency data. This is similar to other observations in separate research [9].

High variance in skill growth trends can be explained by some underlying causes. While the self-assessment and test topics remained consistent over the years, the event content emphasis was re-negotiated by the joint training program leadership team prior to each event to maintain relevance for all participant institutions in the face of real world events. The program leaders' rapid modification of the content was essential to achieve high levels of institutional engagement by addressing changing priorities of stakeholders. This deemphasized some skills tracked in the assessment design. The stakeholders often became aware of the new priorities upon review of tabletop and other exercises.

The need for additional types of research became apparent upon review of the data in relation to the events. The PELE approach used by the participating institutions between training events was significant, allowing the individual capabilities to rise between events even though post-surveys did not suggest strong technical advancement in single-weekend events. This is reflected in the significant growth from a longitudinal view. Creating a coordinated PELE should enhance the various agencies ability to both consistently cover identified areas and make agile moves in a more structured way. This research did not cover the PELE systems or processes but merely observed the need to track inter-event progress of participants to fairly attribute the source of skill growth. The authors confirmed this in after-action briefings with joint training program leadership. Additionally, there is a need for more detailed study of each component of this general program to determine what value each is contributing. A behavioral study assessing the specific causes of learning outcomes could lead to better rebalancing the portfolio of activities. Finally, this study did identify significant growth in the participants of the program, particularly over the longitudinal span of years. The general method presented could address skill gaps seen in the measured areas for a range of similar collaborative ventures. Future studies of this method would require better tracking of growth related activity or specific behavioral studies of learning activities.

Citations:

1. Kenney, M Cyber-Terrorism in a Post-Stuxnet World, Orbis, Volume 59, Issue 1, Pages 111-128 (2015)

2. Novak H, Moore, E., & Likarish, D.. A Cyber Security Multi Agency, Collaboration for Rapid Response that Uses AGILE Methods on an Education Infrastructure. In Information Security Education Across the Curriculum, pp. 41-50. Springer International Publishing (2015)
3. Saunders, J. H.. Simulation approaches in information security education. In *Proc. 6th National Colloquium for Information System Security Education, Redmond, WA*. June (2002)
4. Perry, R. W.. Disaster exercise outcomes for professional emergency personnel and citizen volunteers. *Journal of Contingencies and Crisis Management*, *12*(2), 64-75 (2004)
5. Petress, K.). What is meant by" active learning?". *Education*, *128*(4), 566 (2008
6. Cheung, R. S., Cohen, J. P., Lo, H. Z., & Elia, F.. Challenge based learning in cybersecurity education. In *Proceedings of the 2011 International Conference on Security & Management* (Vol. 1), July (2011)
7. Bartel, A. P.. Measuring the Employer's Return on Investments in Training: Evidence from the Literature. *Industrial Relations*, *39*(3), 502 (2000)
8. Creswell, J. W., & Clark, V. L. P.. Designing and conducting mixed methods research, SAGE Publications Inc. (2007)
9. Eckerson, W. W.). *Performance dashboards: measuring, monitoring, and managing your business*. John Wiley & Sons (2010)