Assessing Performance in an Innovative Cybersecurity Pilot Course

Patrick Shane Gallagher, Ph.D.
Institute for Defense Analyses
Alexandria, VA
pgallagh@ida.org

ABSTRACT

In 2014, 25% of all organizations polled across industry said the lack of infosec skills were a problem. In 2015, an Enterprise Strategy Group (ESG) survey found that 28% reported a shortage of infosec skills (Trendmicro, 2015). With the growing threat of cybercrime and national security issues, growing the number of qualified cybersecurity professionals has become a national imperative. As the cybersecurity universe is shaped by new technologies, unknown threats, and increasing vulnerability in a dynamic environment, there is an established need to rapidly establish innovative, effective, efficient and responsive cybersecurity education initiatives (Dark & Mirkovic, 2015). One such initiative recently piloted by the Department of Defense is the Cyber Operations Academy Course (COAC). The first pilot began in May 2015 at the Fort McNair campus in Washington D.C. As a six-month immersive course, participants consisted of 20 mostly military personnel from all four branches of the military services, various backgrounds and little if any cyber experience. Employing an authentic problem-based course using cooperative and collaborative learning models, the pilot consisted of instruction in foundations, defensive/offensive operations, programming, social engineering, and skills integration. Leveraging cyber ranges and capture the flag (CTF) activities, the course was also supported by four "fireteam" leads as facilitators, coaches, and subject matter experts. At the end of the course, students developed cyber capabilities and tools, developed and deployed exploits, detected and responded to incidents, and used social engineering to exploit "targets." In comparison with existing cyber protection teams deployed in DoD installations, the students were as capable and in some cases more capable in comparisons of performance. In pre/post comparisons, students exhibited potentially large knowledge gains. This paper discusses the nature of the course's pedagogy; the challenge of developing representations of learning outcomes and performance; and the challenges in developing performance-based assessments to authentically and objectively assess students' knowledge and skills in the context of the course

ABOUT THE AUTHOR

Dr. P. Shane Gallagher is employed by the Institute for Defense Analysis and is supporting the Advanced Distributed Learning (ADL) Initiative and OUSD Force Training as a learning scientist and education specialist. He received his Ph.D. in Instructional Technology from George Mason University and MA in Educational Technology from the University of New Mexico. Currently, Dr. Gallagher provides learning science and methodological direction for applied research projects and cybersecurity assessment and is the lead researcher for assessing the development of the ADL Total Learning Architecture. Dr. Gallagher has directed research on video game design for cognitive adaptability and learning science implications of the design of the xAPI and is also researching methods to apply the xAPI and its syntax to describe social learning interactions and human performance especially within cyber-physical contexts. He has led research projects in cognition and game design and R&D projects in learning object content models, simulations, reusable pedagogical models, organizational readiness, and knowledge management. He has been recognized by NASA for his work on assessing the Johnson Space Center on knowledge management readiness by the JSC Chief Knowledge Officer and has authored papers and chapters on neuroscience, cognition, game design, and innovative learning technology applications and specifications.

Assessing Performance in an Innovative Cybersecurity Pilot Course

Patrick Shane Gallagher, Ph.D.
Institute for Defense Analyses
Alexandria, VA
pgallagh@ida.org

BACKGROUND

Cyber warfare, cyberterrorism, and cybercrime are serious existential threats to the national security of the United States. With all of the high profile cybersecurity breaches that have occurred in the private sector (e.g. Target and Home Depot) as well as government (i.e., Office of Personnel Management and the Joint Chiefs of Staff), this issue is now affecting everyone throughout society (DiGiovanni, 2015). Directly addressing these threats is the creation of a Department of Defense (DoD) Cyber Mission Force forecasted to be fully manned and fielded by the end of 2018. However, fully manning the Force requires increasing the number of trained cyber operators and their accession to meet Strategic Goal I: Build and Maintain Ready Forces and Capabilities to Conduct Cyberspace Operations (Carter, 2015).

There simply aren't enough trained personnel to counter the myriad numbers and potential scale of cyber-attacks from determined adversaries, and the need for increasing the number of qualified cybersecurity professionals throughout all sectors has now become a national imperative (DiGiovanni, 2015). Within the private sector, in 2014, 25% of all organizations surveyed said a lack of infosec skills were a problem. Additionally, in 2015, an Enterprise Strategy Group survey found that 28% reported a shortage of infosec skills (Trend Micro, 2015). As the cybersecurity universe is shaped by new technologies, unknown threats, and increasing vulnerability in a dynamic environment, there is a pressing need to rapidly establish innovative, effective, efficient and responsive cybersecurity education initiatives (Dark & Mirkovic, 2015).

Members of the Cyber Mission Force are military cyber practitioners or cyberwarriors and as such will require training in the foundational skills to execute each of the three Cyber Mission Force missions. To meet this need, the Services modified existing training courses in information technologies, intelligence, and communications to train military cyber practitioners utilizing a traditional, formal education learning environment. In 2012, These courses evolved into a tri-service curriculum called the Joint Cyber Analysis Course (JCAC) which was developed by the National Security Agency (NSA) (DiGiovanni, 2015). Those identified for attendance to JCAC usually comprise graduates of various programs from the service academies and schoolhouses (Li & Daugherty, 2015).

JCAC is centrally located within the cybersecurity training pipeline. It functions generally as the gateway between service specific training and either DoD or service cyber mission teams or, ultimately, for further training and induction to the NSA. Although centrally important to the training pipeline, JCAC isn't the only avenue for accession with the NSA. There are multiple ways currently used to identify prospects. Two of the primary alternate avenues are referrals from colleges including institutions funded as Centers of Academic Excellence (CAE) by the NSA and Department of Homeland Security (DHS) and identification through competitions (such as CTFs) and outreach activities (Li & Daugherty, 2015). This pipeline is evolving but currently resembles the illustration in Figure 1.

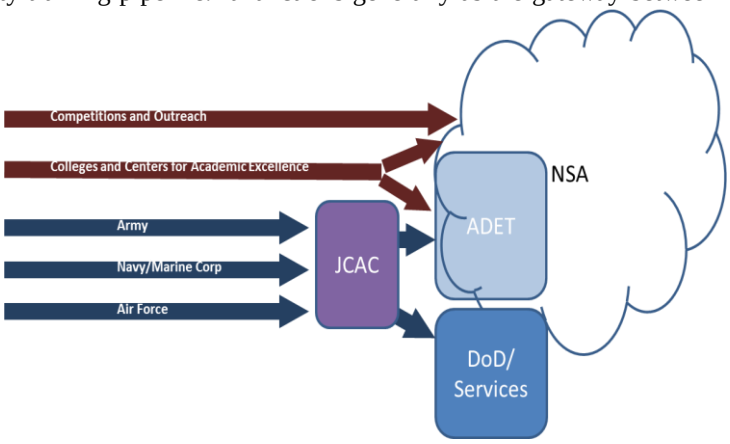

Figure 1 Cybersecurity Pipeline

JCAC was in place to prepare warfighters in a comprehensive manner for further specialization who are considered mission ready to participate in various service Cyber Mission Force Teams. However, it was discovered through an evaluation exercise led by the director of the DoD Operational Test and Evaluation (OT&E) activity and observed by the Director of DoD Force Training (FT[1]) that teams' proficiency had high variability on foundational knowledge and skills and they couldn't problem solve as a team, leading to the conclusion that prior training was lacking in developing these outcomes.

After a review of JCAC curriculum was performed, it was determined that although it was appropriately aligned to the operational needs of the mission tasks, it taught in a traditional "transmission" model. Instead of team based projects and problem solving, it focused on declarative knowledge acquisition providing limited hands on training in the cyber skills. What hands-on training existed consisted of the use of scripts in the classroom that walked the student through a set of checklist-based procedures in response to cyber problem sets (DiGiovanni, 2015).

In response to an observed deficiency in the training methods for cyber operators, the FT Directorate created and piloted a course called the Cyber Operators Academy Course (COAC), with the first pilot occurring May through October 2015. Based on similar learning outcomes to JCAC, the course design originated with the Director of FT. He wanted to use largely a journeyman-apprentice learning model incorporating situated learning, problem-based learning, experiential learning, and cognitive apprenticeship (F. DiGiovanni, personal communication, 2016). This desire situated the pilot within a constructivist epistemology prevalent in the learning science literature by such theorists as Vygotsky, von Glasersfeld, Dewey and others (Dewey, 1938; Glasersfeld, 1995; Vygotsky, 1978) and highlighted by the National Research Council (Bransford, Brown, & Rodney Cocking, 2000; Pellegrino & Hilton, 2012). These theories were to be embodied in the curriculum through the structure of the course, the incorporation of "Fireteam leads," and the dynamic and relevant nature of the topics and related problems (F. DiGiovanni, personal communication, 2016).

CYBER OPERATORS ANALYSIS COURSE

During the initial pilot, the course's working title was CyberCore. Subsequently, the name evolved to the Cyber Operators Analysis Course or COAC. In COAC, students were divided into learning or "fireteams" based on the infantry's tactical organizational structure (Nugent, 2006) and the U. S. Army's Warrior Leaders Course (U. S. Army, 2016). Fireteams worked together collaboratively and cooperatively to bond as a team, solve assigned problems, and compete with each other and in external events. Fireteam leads were assigned to each fireteam and provided mentoring, scaffolding, direction and motivation. They generally acted together as the course's overall instructors. The fireteam leads were also highly qualified subject matter experts in the domain of cybersecurity previously heavily involved working with and for government intelligence agencies in offensive and defensive cybersecurity activities.

Each day would begin at 0800 with a large group meeting, where either a problem of the day (or much longer time period) was presented or discussion and insight about the current assigned problem occurred. Around 0815 to 0830 the class was dismissed to the respective fireteams and team rooms to work on the assigned problem. Problems ranged from such things as building the team server from the ground up including the software stacks, reverse engineering, malware identification and removing, creating key loggers, or creating exploits. In most cases students had little to no experience in doing any of them. The class would come together again at noon to sync up on progress, discuss common issues and needs and receive further insight or instruction. This lasted for another 15-30 minutes after which they would disperse to their team rooms to continue working. Fireteam leads were consistently monitoring their teams as they were working to provide necessary scaffolding and motivation or direct instruction about a particularly hard concept. In addition, a course-long learning activity on social engineering was assigned with the final results presented in a portfolio review at the end of the course.

Supplemental learning activities were also scheduled throughout the course. For example, nationally and internally recognized experts in various cyber operations and tools would guest lecture and/or provide or guide hands-on activities. This included such topics as socio-anthropological issues, special tools, lock picking, or social

[1] Office of the Assistant Secretary of Defense (Readiness)

engineering. Other activities consisted of capture theflag-type events with competition both against each internal team as well as against external teams.

The physical learning environment for COAC was at Ft. McNair next to the Inter-American Defense College (IADC). Those facilities had accommodations for the larger group meetings as well as small team working spaces (dorm rooms). For team working spaces, teams were issued their own dorm room consisting of two large rooms and a bathroom each. Teams set up in either in the first large room with a table, the rear room, or both to work. A common area on each floor was available with kitchen facilities including a refrigerator and coffee machine. This environment provided a sequestered work space for each team to work in as long as they wanted with some students working easily into the night or over weekends on their own. Students were provided with workstations for individual work as well as the components to build and setup their own team server. Each team room server was networked with students' workstations with the team room network connecting to the Internet over wifi access points to a commercial Internet drop.

Course Outcomes

There were no explicitly stated learning outcomes other than the list of learning objectives provided in the evolving course documentation. After performing cognitive task analyses of the problems assigned to the students using the fireteam leads as subject matter experts, the actual learning outcomes and knowledge domain structure became more apparent. The outcomes centered on three primary areas: offensive/defensive operations, software engineering, and socio-anthro or social engineering. Main problem categories emerged and defined the following high level outcomes: decide whether a breach has occurred and respond appropriately (defensive operations), exploit a buffer overflow (offensive operations), create and deploy a key logger to collect information from a target (software engineering), and develop a portfolio of a human target with positive contact (social engineering). In addition, general cybersecurity knowledge was identified that, although not inherently needed to accomplish tasks in any of the other problem areas, would be generally something that should be known by the cyber operator. The end result of the cognitive task analysis was created as knowledge/learning outcome maps and used to drive the creation of a performance assessment.

Need for Evaluation

Although the course implementer included some pre-testing at the beginning of the course it was minimal and used mainly for attempting to sort students. There was no comprehensive instrument used and therefore no baseline established to understand what, if any, learning gains were made. However, there were various assessments sprinkled throughout the course. One such activity was a capture the flag game called PICO CTF developed by Carnegie Mellon University (Carnegie Mellon University, 2014) used for placing high school student in their cyber program. Others included an industry certification test – i.e., Offensive Security Certified Professional examination (Offensive Security, 2016). These assessments provided some measures that were potentially useful but overall weren't designed to be sensitive to pre/post learning gains or due to intellectual property issues wouldn't provide any student performance data.

There were many learning objectives listed by topic in the course design documentation but there were few if any alignments between them and what was really going on in the classroom. Those listed were also of a lower level and mostly declarative in type. This was most likely due to sponsor's desire for the course to be taught using the journeyman/apprentice model contrasted with the use of traditional instruction design methodologies by the contracted course implementer. This contrast inherently created some tension in the execution of the initial pilot.

In addition to the lack of a comprehensive pre/post assessment methodology and misalignments, there was no inherent evaluation strategy for comparing the pilot to an existing course such as JCAC or other comparable populations. This situation led to the request for the Institute for Defense Analyses to provide an in situ assessment and evaluation of the pilot COAC to attempt to understand if the learning model employed could help bridge the gap identified in the operational mission evaluation of JCAC graduates by OT&E. This resulted in the need to retrofit an assessment and evaluation strategy to the undergoing pilot two months after it had begun.

For any assessment and evaluation strategy the goals need to be determined. In this case, the goals of the COAC pilot evaluation were primarily to determine to the greatest degree possible what student learning gains occurred as a

result of the course. This goal resulted in the following research question: As a result of the course (treatment) have changes in learning occurred? If so what were they? Additionally, and to better understand how the pilot might indeed bridge the gap, the next goal was to attempt to compare students' knowledge and skills at the end of the course with some type of external program or normed data set to determine potential worth. To meet this goal, the following research question was posed: How do the students' performances compare on an individual and group level on common examinations to performances by others?

METHODS

Design and Variables

The evaluation design was a quasi-experimental design using as much existing data as possible integrated with new and more comprehensive data about students' performance, knowledge, and skills. It was decided to use a pre/posttest design to assess gains in knowledge and performance and a between subjects design for evaluating performance with a control group. As the Institute for Defense Analyses (IDA) evaluation didn't begin until over a third of the way into the course and an evaluation strategy hadn't been put in place by the course designers, there wasn't any design for a comprehensive pre-test and matching post-test to assess students' learning gains. Therefore, IDA was forced to use what initial data was taken as a partial baseline and construct a method to attempt to re-create a comprehensive student baseline of students' knowledge and skills levels relative to the complete set of identified learning objectives. To this end, the existing data points taken before the treatment combined with a baseline knowledge evaluation of each student by the fireteam leads comprised the data for pre-treatment knowledge and skills. These are noted as pre-treatment dependent measures $O_{1..3}$ in Figure 2 below. Post-treatment measures are also noted as $O_{1..3}$ corresponding to the pre/post alignment. Several post measures were used to assess subsets of the learning objectives that comprised $O_{3.}$ Specific objectives rated in the Baseline Knowledge evaluation were performance-based and were able to be assessed through two assessments—the Performance Assessment and the OSCP exam. Post treatment, they were also used for the between groups design. Dependent measure O_4 were test results from a digital media collection exam, didn't correspond to specific learning objectives and was used for informative purposes only. Dependent measure O_5 was the output from an all-against-all capture the flag (CTF) used for the between groups design with the team as the unit of analysis and is reserved for future analysis.

Specific dependent variables used for analysis in the study were:
InitialTech for pre/post technical knowledge - O_1
picoCTF for pre/post technical conceptual and applied knowledge – O_2
BaselineKn for comprehensive baseline pre/post rated knowledge and skills – O_3
IndPerfAssmt for Individual Performance Assessment of applied knowledge– O_3
OSCPExam for individual assessment of advanced applied knowledge leading to certification – O_3
IndPerfAssmtTSCR for Individual Performance Assessment total scores – O_3
IndPerfAssmtOBJCOM for Individual Performance Assessment Objectives Completed – O_3

	Initial Tech	PicoCTF	BaselineKn	CyberCore Crs	DCITA EnCase	DCITA Exam	PicoCTF	Ind Perf Assmt	OSCP Exam	Team CTF	Dark Web	Initial Tech	General Kn
N-CyberCore	O_1	O_2	O_3	X	X	O_4	O_2	O_3	O_3	O_5	O_3	O_1	O_3
N-CyberCore	O_1	O_2	O_3	X	X	O_4	O_2	O_3	O_3	O_5	O_3	O_1	O_3
N-CyberCore	O_1	O_2	O_3	X	X	O_4	O_2	O_3	O_3	O_5	O_3	O_1	O_3
N-CyberCore	O_1	O_2	O_3	X	X	O_4	O_2	O_3	O_3	O_5	O_3	O_1	O_3
N-INSCOM								O_3		O_5			
N-INSCOM								O_3		O_5			
N-Army Cyber								O_3		O_5			
N - Coast Guard								O_3		O_5			
INSCOM is 780th MI BDE INSCOM													
Army Cyber is Cyber Protection BDE													
Coast Guard is CG CyberCom													

N=nonequivalent groups
Each group = ~ participants
CyberCore = 1 group for all except O_4
[shaded box] = completed

Figure 2. Evaluation Design

Population

The student population for the COAC pilot consisted of 21 personnel from the U.S. Army, U.S. Air Force, U.S. National Guard, U.S. Navy, and DoD Civilians. All but two had no previous Cyber related experience with most of them having only zero to two years of prior work experience related to information technology (IT). Although nine of the 21 had IT related education experience, most did not have college degrees, and the majority of those from the services were enlisted with ranks of E3 or E4.

For comparative purposes, 18 personnel representing Army Cyber and Coast Guard Cyber commands were obtained. All of them were in cybersecurity related military occupation codes (MOSCs). Most were enlisted with ranks of E6 or E7, did not have college degrees and reported five to six years related work experience.

Primary Instrumentation and Data Collection

There were various instruments collecting data on the students throughout the course. Due to pre-design deficiencies, not all were of value.. As designated in Figure 1, each dependent variable measure is described by O_x. This notation describes the how the instrument was deployed for data collection in contrast to the independent variable notated by the X_x. Instruments with the same subscript indicates that they are assessing the same learning objectives. For example, Baseline Knowledge and Skills captures instructor ratings on every listed learning objective pre-treatment. The Individual Performance Assessment and the OSCP have the same subscript indicating that they are measuring the same or a subset of the same learning objectives. Demographic data were collected on each student consisting of basic biographical data, service data and/or employment data. The following variables were used to assess student performance and evaluate the course potential

InitialTech (11 Item Technical Test) – O_1: To assess primarily declarative technical knowledge, a limited contractor designed knowledge test was used consisting of 11 questions. Students were given this test during the first week of May 2015. It was administered again to both students and control participants during the fourth week of October 2015.

picoCTF – O_2: This was administered at the beginning of the course and was an individual capture the flag game during the first week of May 2015. It was administered again during the first week of October 2015. picoCTF is a game-based cybersecurity problem solving environment with points assigned for challenges of increasing difficulty. Gains were measured on the difference in points from pre to post administration. Students were given the weekend to work through the game which presented issues in validity of measurement. Most challenge solutions were freely available on the Internet and the game really only measured forward progress.

BaselineKn (Baseline Knowledge and Skills Rating) – O_3: To gain a comprehensive baseline of the students' knowledge and skills, a rating method was used. Each fireteam lead was asked to rate all students on a five-point scale against 205 learning objectives validated as within scope of the curriculum. These objectives were comprised of some knowledge level but mostly application-level objectives. Also, a large portion directly corresponded to the OSCP (described below). Measures used were students' average scores across all objectives for what fireteam leads' rating of students'proficiency at the beginning of the course and at the end. Essentially reverse engineering a baseline, the beginning rating was taken closer to the midpoint of the course as the IDA evaluation task didn't actually begin until two months after it started. This required the fireteam leads to think back on the initial knowledge and skills for each student introducing a validity threat. However, as all four fireteam leads rated all students, there was some measure of inter-rater reliability established. Initial ratings occurred during the 9th week of the course or the second week of July 2015. Final ratings occurred during the third week of October 2015.

IndPerfAssmt (Individual Performance Assessment) – O_3: As a post-test only, a performance-based assessment was devised using a beta version of the Project ARES (Circadence, 2016) game-based AI-enabled cybersecurity training environment. Using authentic Linux environments supported by mission tailored virtual networks, the student had to complete three specifically designed missions each with three performance objectives aligning to the performance outcomes of the course. Measures taken were counts of objective completions per mission, mission completions, and time to complete an objective and a mission per student. Total scores were also assigned to each participant which were used for group comparisons. Students and control participants were given a three days to complete three missions during the third week of October 2015.

OSCPExam (Offensive Security Certified Professional (OSCP) Examination) - O_3: To provide the opportunity for leaving the course with a certification, students were asked to take the OSCP Exam. This is an intensive performance-based exam with 24 hours to complete focusing on offensive techniques. Results from the exam are only provided to the individual as a pass/fail. There is no scoring or normed data available making it of limited value for understanding knowledge and skill deficits from the ones taking the exam. Measures used were pass-rate percentages compared to average pass rates of all takers and pass to fail from within the student population. It was informally reported that the general pass-rate of examinees was approximately 20%. There is no documentation but this is the benchmark used for analysis. Students took this exam during the second week of October 2015.

Other Measures: Other measurement and evaluation data were collected throughout the course. One such opportunity was during a three-day event called Cyber Stakes Live hosted by David Brumley of Carnegie Mellon University. Another opportunity was during a week-long event called Tracer Fire hosted by the U. S. Department of Energy. Data from these events are not included at this time and are reserved for further analyses activities.

Timeline

COAC began during the first week of May 2015 and ended the last week in October 2015 for a duration of six months. Students received days off for holidays but were otherwise engaged full time with the course during this period. Pre-testing occurred during the first week of the course with some potentially measureable activities occurring sporadically throughout. Such activities consisted of capture the flag evaluations run by third parties and other government agencies that were leveraged by the course designer (briefly discussed above as Other Measures). Post testing occurred during the third week of October 2015.

ANALYSIS AND RESULTS

Parametric and non-parametric methods were used in the analysis. These methods comprised T-tests for pre/post measure comparisons for $O_{1..3}$ and an analysis of variance (ANOVA) for between group comparisons on O_3 only. Simple counts, scores, and percentages were also used for O_3 measures for the performance-based assessment reported individually and pass-rates were used for the OSCP.

To answer the first research question (as a result of the course (treatment) did changes in learning occur), the following results were produced:

- For InitialTech pre/post there were significant gains produced ($p<.001$) with an effect size = 1.05,
- For picoCTF pre/post, students showed significant gains ($p<.001$) with an effect size of 1.83, and
- For BaselineKn, instructor ratings for 205 learning objectives taken pre and post intervention showed significant gains ($p<.0001$) with an effect size = 4.02. Results are displayed graphically below in Figure 3.

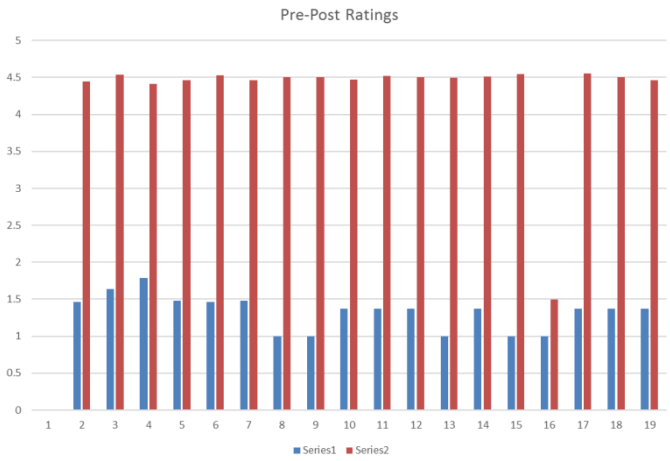

Figure 3 Baseline Knowledge Pre/Post Ratings

To assess IndPerfAssmt the Individual Performance Assessment was used. Counts of completed objectives by mission were taken as well as counts of completed missions (requiring all mission objectives to be completed). Also, time spent on each objective and mission was measured. The unit of analysis for this assessment was at the individual level and was post-treatment only. Due to absenteeism the student number assessed was 16 for the first two missions and 13 for the third. Also taking the assessment were members of the control group. For the first two missions their number was 8 each and 6 for the third

For Mission 1 Disable C&C (command and control) Botnet Server the objectives were:

1. Scan network
2. Brute force login
3. Kill C&C botnet webserver.

The average completion times for each objective were:

1. 27:11
2. 45.42
3. 9:01.

Three out of 16 students complete all objectives and the mission for a 19% student completion rate and a combined rate or 13%. Five out of eight control group students complete all objectives and the mission for a control completion rate of 63% and a combine rate of 33%.

For Mission 2 Steal Bank Account Information the objectives were:

1. Identify target via scan
2. SQL injection
3. Exfiltrate data.

The average completion times for each objective were:

1. 11:43
2. N/A
3. N/A

None out of 18 students completed the objectives and the mission for a 0% completion rate.

For Mission 3 Steal Mission Documents from Airbase the objectives were:

1. Find vulnerable service
2. Deliver exploit
3. Retrieve file

The average completion times for each objective were:

1. 3:16:04
2. 1:14
3. 23:05

Seven out of 13 students completed all the objectives and the mission, for a student and combined completion rate of 46%. There were no mission completions by the control group. A graph of student performance on Mission 3 is included as Figure 4. Six participants were from the control group. If no bar appears on the graph in Figure 3 beside a user ID, then the participant didn't complete the first objective but logged on to the system. Graphs for the other missions are displayed in a similar fashion.

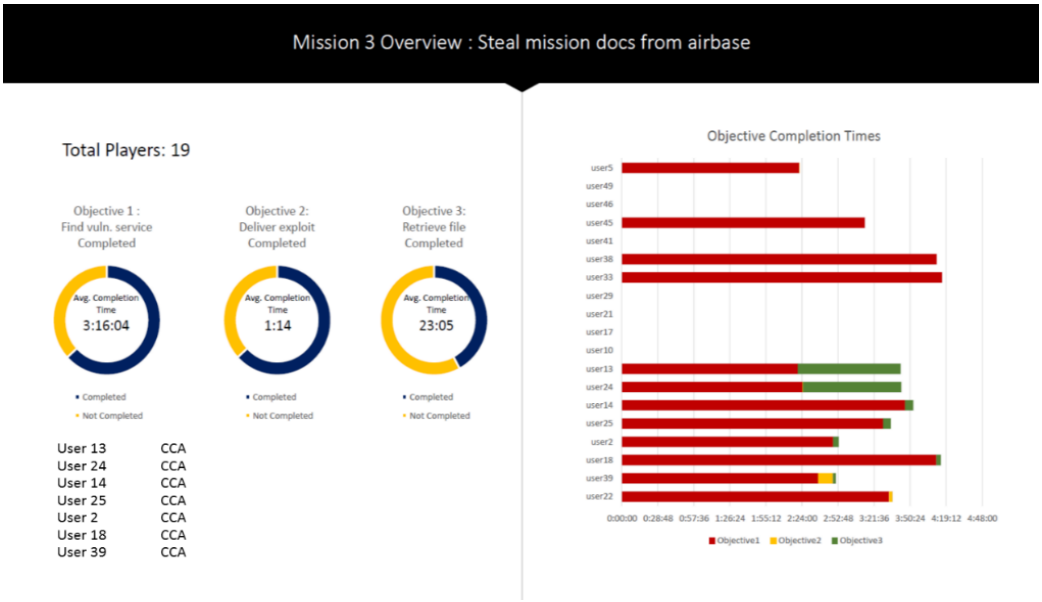

Figure 4 Mission 3

Further assessment of the learning objectives tied to O_3 directly aligned to the Offensive Security knowledge and skills was the OSCP examination assessing OSCPExam. This also functioned as an estimated measure against a greater population of cybersecurity professionals. There were 12 attempts out of 21 students to take the OSCP Exam. Six students passed and six students failed. This produced a 50% pass-rate far exceeding the informal norm of 20% of those taking the examination.

To answer the research question "how do the students' performances compare on an individual and group level on common examinations to performances by others", comparisons were made to the control group on the 11 Item Technical Questionnaire (O_1), the total scores of the Individual Performance Assessment, and the number of objectives completed by mission on the Individual Performance Assessment (O_3). The following results were produced:

- For InitialTech, there was no significant difference between groups ($p>.3$)
- For IndPerfAssmtTSCR using Individual Performance Assessment total scores, there was no significant differences between groups ($p>.9$)
- For IndPerfAssmtOBJCOM using Individual Performance Assessment Objectives Completed there was no difference between groups on Mission 1 (Kill botnet server) and Mission 2 (Steal act. Info)
- For For IndPerfAssmtOBJCOM there was a significant positive difference ($p=.002$) on Mission 3 (Exploit server for information - offensive) between the students and the control groups.

DISCUSSION AND CONCLUSION

After six months of full time immersion it was apparent that significant learning gains occurred in the students' general technical knowledge and skills in the primary knowledge domains of the course. What isn't known is the true breadth and depth of those gains. The most significant increase came from the pre and post ratings of the fireteam leads. Although insightful, this does not represent objective measures and could be quite biased due to the time spent working together over the course duration. What may have more validity but little breadth was the 11-item knowledge test. Students showed significant gains with a measurable effect size. The problem with this is the limitation of the test itself as well as a potential learning effect. Still, this did produce somewhat objective results. The gains on the Pico CTF, although significant, are difficult to attribute to the course. It suffers from the potential of a large learning effect as well as the ability for challenge solutions to be found online. It was reported anecdotally that simply searching for the solutions was quite common among the students.

Performance on the Individual Performance Assessment was perhaps the most interesting. All challenges represented a problem type but not an exact problem they had experienced in class. The students did not fare well on the first two missions but did well on the third and hardest mission. They had been working on offensive tactics and exploits since the beginning of the course and Mission 3 was the most similar to that. This speaks to their ability in very similar problems but not too well in transference.

What might be the most interesting however, is the comparison to external participants already functioning as cyber operations teams. Performance on measures of technical knowledge and defensive cybersecurity performance tasks showed no differences between the students and the active cybersecurity military professionals comprising the control group. It would be expected that the control group would perform significantly better in all aspects as they had been through JCAC and are considered mission ready especially with defensive tactics.

Of great interest but not unexpected is that performance on measures of offensive cybersecurity performance tasks showed significant increase in performance by the students over the control group. This first COAC pilot had an appreciable offensive lean to the learning in its approach, problem sets, and general mindset. Defense was not emphasized as much and the students may have suffered some because of it. However, the philosophy was one of understanding how and why opponents attack in order to understand how to better defend against it. It is well known that cyber protection teams are defensive focused and this was apparent in the scoring on the Individual Performance Assessment.

Passing the OSCP is difficult for most cyber practitioners. It was not unexpected that only half of the students took the exam. However, a 50% pass-rate for those that did take it was beyond the rumored norm of all examinees. As

this is an internationally recognized test, this speaks to the level of proficiency that some took away from the course compared to an international population of potential cyber operators.

In conclusion, the course produced learning gains in the students compared to their initial knowledge and skills coming into it. It is unclear if this is due to the pedagogical model embedded within the course but learning science points to this type of model as being optimal (Gallagher, 2013). In comparing the students' performance with others it is apparent that they performed as well or better than current mission ready cyber operators. What's interesting about this is that although no baseline performance data could be obtained from JCAC students for this study, using the control group as a proxy produced very favorable results as something either comparable or potentially better.

The results point to the need for further study and another pilot with an embedded assessment and evaluation strategy. With that in mind, it may also imply some inherent changes in the way training occurs within the cybersecurity pipeline. These changes should include a migration away from traditional models of instruction that emphasize the learning and retention of facts and processes without the conceptual understanding that comes with experiential learning within real problem contexts. Anecdotally, students from the first COAC pilot that had also taken JCAC expressed that after participating in COAC they finally understood what was presented to them in JCAC. Although not present in the quantitative data discussed in this paper, these types of comments drive right to the heart of why the current format of most existing cybersecurity pre-NSA training programs might not be effective.

Also, not explicitly discussed previously in this paper, during a final in-class 48 hour CTF using existing service teams as a control, a surviving COAC team came in second to an existing mission-experienced team from INSCOM (U.S. Army Intelligence & Security Command). Also of interest, another INSCOM team packed up during the night and left in frustration because they hadn't been taught or experienced the offensive techniques needed to win or even place. The latter also implies that emphasis on offensive techniques as lacking in current service training. In employing a problem-based situated learning environment, it could be beneficial for offensive problems and skills to be incorporated. However, when analyzing the current NICE (National Initiative for Cybersecurity Education) competency framework currently being tailored by the DoD (Li & Daugherty, 2015) and the programs in the colleges and CAE's, it is also apparent that little if any attention is placed on offensive cybersecurity techniques. This could point to the potential need for systemic realignment.

As most of the students involved in the first pilot hadn't even experienced any service training necessary to be admitted to JCAC given their high level of performance individually and in teams against existing mission cyber teams, there are implications that the pipeline itself may need some rethinking. Learning to be a cybersecurity operator requires problem solving ability, perseverance, motivation, and passion, and is not something that can be adequately trained in a linear traditional model. Also, personnel who probably are best are the ones that might least appear to be traditional in the sense of a warfighter (J. Rigney, primary COAC fireteam lead and subject matter expert, personal communication, 2016). This implies that at the very least there could be a place for the type of learning that COAC embodies whether it is a course in place of or additional to JCAC. There are also implications that traditional models of service accession may not be appropriate for the type of personnel needed to by a cyberwarrior. Much like the way the Air Force evolved from the Army Air Corps, it might be that the only real way to access and train the folks needed for cyberwarriors is to create a completely new service – Cyber Corps.

ACKNOWLEDGEMENTS

The author would like to acknowledge Mr. Frank DiGiovanni for the intellectual and strategic thinking driving key concepts in this paper and for providing the opportunity to fulfill the assessment needs of the Cyber Operators Academy Course.

REFERENCES

Bransford, J., Brown, A., & Rodney Cocking (Eds.). (2000). *How People Learn: Brain, Mind, Experience, and School*. Washington D.C.: National Academy Press.

Carnegie Mellon University. (2014). PICO CTF 2014. Retrieved from https://picoctf.com/about

Carter, A. (2015). *The DoD Cyber Strategy*. Washington D.C. Retrieved from http://www.defense.gov/News/Special-Reports/0415_Cyber-Strategy

Circadence. (2016). Project ARES Overview. Retrieved from https://www.circadence.com/project-ares/overview/

Dark, M., & Mirkovic, J. (2015). Evaluation theory and practice applied to cybersecurity education. *IEEE Security and Privacy*, *13*(2), 75–80. http://doi.org/10.1109/MSP.2015.27

Dewey, J. (1938). *Experience and Education*. New York: Kappa Delta Pi.

DiGiovanni, F. (2015). The Value of the Journeyman-Apprentice Learning Model for Military Cyber Practitioners. Philadelphia: University of Pennsylvania.

Gallagher, P. S. (2013). Transforming Educattion through Neuroscience, Cognition, and Game Design. In M. Davies (Ed.), *Proceedings of the 3rd International Transformational Leadership Conference*. Washington D.C.: National Defense University.

Glasersfeld, E. von. (1995). *Radical Constructivism*. New York: Routledge Falmer.

Li, J. J., & Daugherty, L. (2015). *Training cyber warriors: what can be learned from defense language training?* Santa Monica: Rand Corporation. Retrieved from http://www.rand.org/content/dam/rand/pubs/research_reports/RR400/RR476/RAND_RR476.pdf

Nugent, F. M. P. J. (2006). *Making a Better Fire Team Leader*. Quantico.

Offensive Security. (2016). Offensive Security Certified Professional. Retrieved from https://www.offensive-security.com/information-security-certifications/oscp-offensive-security-certified-professional/

Pellegrino, J. W., & Hilton, M. L. (2012). *Education for life and work: developing transferable knowledge and skills in the 21st century*. *National Academies Press*. http://doi.org/0-309-25649-6

Trend Micro. (2015). The challenges of cyber security education and training in 2015. Retrieved from http://blog.trendmicro.com/the-challenges-of-cyber-security-education-and-training-in-2015/

U. S. Army. (2016). Warrior Leader Course.

Vygotsky, L. S. (1978). *Mind in Society*. (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge: Harvard University Press.

APPENDIX B ACRONYM LIST

ADET	U.S. NSA Associate Directorate for Education and Training
ANOVA	Analysis of variance
CCA	Cybercore Academy (students from first pilot)
CAE	Center for Academic Excellence
COAC	Cyber Operators Academy Course
CTF	Capture the Flag
DoD	U.S. Department of Defense
DHS	U.S. Department of Homeland Security
FT	Force Training Directorate
IADC	Inter-American Defense College
ID	Identification
IDA	Institute for Defense Analyses
INSCOM	U.S. Army Intelligence & Security Command
IT	Information Technology
JCAC	Joint Cyber Analysis Course
MOSC	Military Occupational Specialty Code
NICE	National Initiative for Cybersecurity Education
NSA	U.S. National Security Agency
OSCP	Offensive Security Certified Professional
OT&E	Operational Test and Evaluation
PICO CTF	Pico Capture the Flag
SQL	Structured query language