Advise When Ready for Game Plan: Adaptive Training for JTACs

Matthew D. Marraffino[1], Cheryl I. Johnson[1], Daphne E. Whitmer[2], Natalie B. Steinhauser[1], & MSgt Aadam Clement[3]

[1]Naval Air Warfare Center Training Systems Division, Orlando, FL

[2]Zenetex, LLC, Orlando, FL

[3]Expeditionary Warfare Training Group Pacific, Coronado, CA

ABSTRACT

The U.S. Marine Corps (USMC) *Commandant's Planning Guidance* (CPG; Berger, 2019) describes a future in which Marines must be prepared to make difficult decisions in high-stress and high-stakes situations. As a result, the *USMC CPG* highlights the need for more flexible, learner-centered training that emphasizes decision-making and critical-thinking skills. One training solution that may fulfill this need is adaptive training (AT), which is training that is customized to an individual's strengths and weaknesses. AT has been demonstrated to be more effective than traditional training methods in some domains (e.g., Landsberg et al., 2011; McCarthy, 2008), however, more research is needed to determine the most effective methods for incorporating AT for complex military decision-making tasks. To these ends, this paper discusses the development of a testbed called Adaptive Training for Terminal Air Controllers (ATTAC) and the results of a training effectiveness experiment.

ATTAC trains game plan development, which is a critical decision-making task that sets the stage for executing close air support (CAS) missions performed by the Joint Terminal Attack Controller (JTAC). Informed by science of learning principles, ATTAC works by presenting a trainee with a series of CAS scenarios, and based on the trainee's responses, the system provides tailored feedback and adjusts the difficulty of subsequent scenarios. Therefore, each trainee receives a unique training experience optimized to one's ability level. In our experiment, we compared the learning outcomes of three conditions: training with ATTAC, training with a non-adaptive version of ATTAC, and a minimal training control.

Our initial results with 52 students in a USMC course showed significantly higher gain scores for those trained with ATTAC compared to both the non-adaptive and control conditions. There was no difference in learning gains between the non-adaptive and control conditions. Implications and lessons learned are discussed.

ABOUT THE AUTHORS

Matthew D. Marraffino is a Senior Research Psychologist in the Simulation and Training to Improve Knowledge and Effectiveness (STRIKE) Lab at NAWCTSD in Orlando, FL. He received his Ph.D. in Human Factors and Experimental Psychology at the University of Central Florida. For almost 10 years, Dr. Marraffino has researched how to apply cognitive theories of learning (e.g., Cognitive Load Theory and Embodied Cognition) to guide the design of trainers using state-of-the-art technology including adaptive training, tablet-based computers, and virtual reality. In addition to publishing his research in a variety of outlets, Dr. Marraffino was nominated in 2017 and 2018 for the Presidential Early Career Award for Scientists and Engineers for his work in applying adaptive training techniques to complex military domains.

Cheryl I. Johnson is a Senior Research Psychologist in the STRIKE Lab at NAWCTSD in Orlando, FL, performing research on emerging training technology and adaptive training systems. She earned her M.A. and Ph.D. in Cognition, Perception, and Cognitive Neuroscience from the University of California Santa Barbara. Dr. Johnson has over 13

years of experience performing technology-based training research. Her research interests include adaptive training, instructional strategies, virtual reality training applications, and multimedia learning.

Daphne E. Whitmer is a Senior Research Psychologist for Zenetex, LLC in support of the STRIKE lab at NAWCTSD in Orlando, FL. She received her Ph.D. in Human Factors and Cognitive Psychology from the University of Central Florida. Dr. Whitmer was a recipient of the 2018 RADM Fred Lewis Postgraduate I/ITSEC scholarship. Her research interests include decision-making, mental models, and adaptive training.

Natalie B. Steinhauser is a Senior Research Psychologist at NAWCTSD in the Basic and Applied Training and Technology for Learning and Evaluation (BATTLE) laboratory. She has 14 years of Navy experience conducting training effectiveness evaluations, developing adaptive training and intelligent tutoring systems, and conducting decision-making research. She holds a M.S. in Modeling and Simulation from the University of Central Florida, and a B.S. in Human Factors/Applied Psychology from Embry-Riddle Aeronautical University.

MSgt Aadam Clement is a USMC JTAC Instructor/Evaluator with over 19 years of Surface to Surface Fires and Close Air Support Experience. He was designated as a JTAC in July of 2008 and has multiple combat deployments to both Iraq and Afghanistan controlling fixed and rotary wing aircraft in support of Special Operations, Joint and Coalition Forces. He is currently the Joint Fires Observer Course Manager and TACAIR Department Staff Noncommissioned Officer in Charge, where he has trained and certified over 500 JFOs and 300 JTACS from multiple services since 2016.

Advise When Ready for Game Plan: Adaptive Training for JTACs

Matthew D. Marraffino[1], Cheryl I. Johnson[1], Daphne E. Whitmer[2], Natalie B. Steinhauser[1], & MSgt Aadam Clement[3]

[1]Naval Air Warfare Center Training Systems Division, Orlando, FL
[2]Zenetex, LLC, Orlando, FL
[3]Expeditionary Warfare Training Group Pacific, Coronado, CA

INTRODUCTION

The U.S. Marine Corps (USMC) *Commandant's Planning Guidance* (CPG; Berger, 2019) describes a future of increasingly complex and volatile environments in which Marines will operate. Such an environment requires Marines to be prepared to make difficult decisions in high-stress and high-stakes situations. As a result, the *USMC CPG* highlights the need for more flexible, learner-centered training that emphasizes decision-making and critical-thinking skills. One training solution that may fulfill this need is adaptive training (AT). Although many definitions of AT exist in the literature, we define AT as "training interventions whose content can be tailored to an individual learner's aptitudes, learning preferences, or styles prior to training and that can be adjusted, either in real time or at the end of a training session, to reflect the learner's on-task performance" (Landsberg, Van Buskirk, Astwood, Mercado, & Aakre, 2011, p. 9). AT has been demonstrated to be more effective than traditional training methods in some domains (e.g., Landsberg, Mercado, Van Buskirk, Lineberry, & Steinhauser, 2012; McCarthy, 2008), however, more research is needed to determine the most effective methods for incorporating AT for complex military decision-making tasks. Therefore, our goal was to perform an experiment to determine whether previously successful AT approaches also apply to a complex decision-making task.

In this research, we focus on the Joint Terminal Attack Controller (JTAC), a certified Service member who is responsible for directing combat aircraft engaged in close air support (CAS) operations. CAS involves taking air actions against hostile targets that are in close proximity to friendly forces. In the following sections, we will describe AT and discuss the theoretical motivations behind a training testbed called Adaptive Training for Terminal Attack Controllers (ATTAC) that was used to test the efficacy of adapting scenario difficulty and feedback to training decision-making skills.

ADAPTIVE TRAINING AND RESEARCH APPROACH

Although one-on-one tutoring is often considered a best practice for training (Bloom, 1984), it is not always a reasonable real-world solution given funding, manpower, and time constraints. Therefore, instructional designers try to approximate this experience through technology-based training solutions, such as AT. Many studies have examined the effectiveness of AT (Billings, 2012; Conati & Manske, 2009; Kelley, 1969; Landsberg et al., 2012; McCarthy, 2008; Serge, Priest, Durlach, & Johnson, 2013), but additional research is needed to examine AT solutions for different task domains. For example, in Durlach and Ray's (2011) review of the AT literature, they were unable to draw strong conclusions about the conditions under which adaptive techniques are most effective because there were few examples of carefully controlled experiments (e.g., few comparisons between an adaptive group and a non-adaptive group or two adaptive groups).

In an exemplary AT experiment, Landsberg and colleagues (2012) explored the effectiveness of AT for training angle on the bow (AOB) in a periscope operator task (i.e., visual-spatial task). In this study, the authors compared trainees' learning gains after training using either an adaptive or a non-adaptive version of their periscope trainer. In the adaptive trainer, the type of feedback and the difficulty of subsequent scenarios was adjusted based on the accuracy of a trainee's AOB (i.e., absolute difference value between the trainee's response and known truth for angle and range). In the non-adaptive trainer, trainees received the same type of feedback and scenario difficulty (that progressed over time) regardless of how they performed during training. Their results showed that students who used the adaptive trainer made periscope calls faster and their training was more efficient relative to those who used the non-adaptive trainer. The goal of the current research was to determine whether these strategies used by Landsberg and colleagues (2012) would also apply in a new context -- a complex, military decision-making task. Therefore, modeling the AT techniques used in Landsberg et al. (2012), we developed the ATTAC testbed to evaluate the effectiveness of adapting scenario difficulty and feedback for a challenging, decision-making task conducted by JTACs during CAS missions.

DEVELOPMENT OF ATTAC

Testbed and Training Task

The ATTAC testbed is a scenario-based trainer created with lightweight software available on most computers. This approach allowed us to be flexible and rapidly develop new capabilities based on input from subject matter experts (SMEs) and instructors. This approach also enabled us to produce an adaptive and non-adaptive version of the trainer for the experiment.

Given the complexities and multi-faceted nature of CAS, we restricted the scope of ATTAC to train game plan development in order to test the effectiveness of AT in the decision-making domain. JTAC instructors identified game plan development as one of the most challenging concepts for JTAC trainees to grasp. The game plan sets the stage for a CAS mission and involves four interdependent decision elements that are passed from the JTAC to the attacking aircraft. The elements include the Type of Attack, Method of Attack, Ordnance, and Interval (TMOI). The Type of Attack (1, 2 or 3) is a requirement set by the JTAC indicating how much control they need over the attack. Method of Attack sets the correlation requirements for how the attacking aircraft and JTAC confirm they are referencing the same target. Ordnance refers to which weapon will be used to prosecute the target. Ordnance selection is based on a number of factors, including which weapon is the best match against a target, concerns about collateral damage, etc. Finally, the Interval determines how much time separation the JTAC requires between attacks for a section of aircraft.

In ATTAC, trainees are presented with text-based CAS scenarios with information required to generate a game plan that would be available to a JTAC in a real-world situation. As seen in Figure 1, each scenario includes Commander's intent, which dictates what effect (e.g., suppress, neutralize, or destroy) the mission is required to achieve. Included below the Commander's intent is a short description of the situation to provide context to the mission, a more detailed description of the targets, along with distances from the targets to the JTAC and nearest friendlies. Each scenario also describes what tools are available to the JTAC (e.g., map and compass, laser target designator, portable laser rangefinder). Below the target description is the aircraft check-in, which is presented in a similar format that a pilot would use to communicate with the JTAC. The check-in includes the quantity and type of aircraft available as well as the weapons and sensor capabilities available.

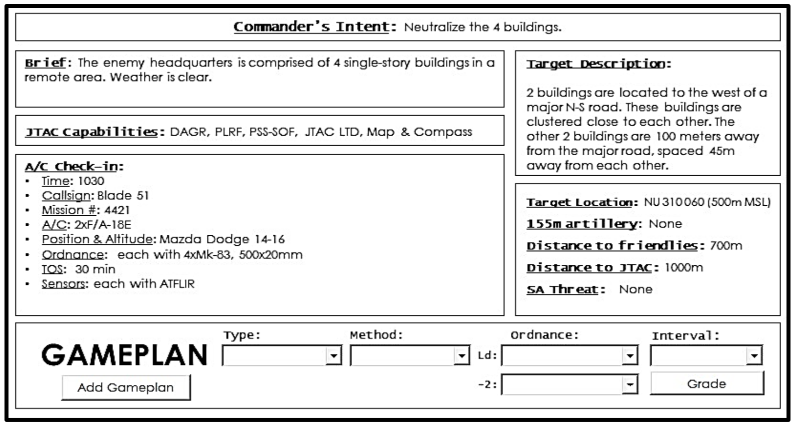

Figure 1. ATTAC's Scenario Interface

Using this information, the trainee selects a game plan using dropdown menus for Type, Method, Ordnance, and Interval at the bottom of the screen. Based on the trainee's answer, the ATTAC system grades the trainee's answer, which ATTAC uses to determine the scenario feedback and difficulty adaptation. After early discussions with SMEs, two things became clear. First, in most cases, several different game plans could meet Commander's intent in a given scenario. Second, the TMOI components are interdependent decisions that interact and affect the overall outcome of the game plan (e.g., a Bomb on Target [BOT] attack may be appropriate for one ordnance, but not effective for another). Therefore, ATTAC was designed to account for and assess multiple TMOI answer combinations and each combination is assessed holistically. The game plan combinations were assigned to three categories based on doctrine and SME input: ideal, acceptable, or unacceptable. The quality assessment of each game plan was based on whether the game plan would meet Commander's intent and whether it was the most efficient, effective, and safe way to accomplish the mission given the information provided in the scenario. A game plan was considered ideal if it was the best game plan to prosecute the target to meet Commander's intent while being efficient and safe. An acceptable game plan was one that may have met Commander's intent, but one or more elements precluded it from being more efficient. Unacceptable game plans contained elements that would prevent mission success or were unsafe for the attacking aircraft or friendly personnel. Using these criteria, we were able to account for the many different ways a JTAC could consider prosecuting the targets in the scenario. Next, we will discuss the theoretical motivations behind the adaptive techniques within ATTAC.

Theoretical Motivation

Cognitive Load Theory (CLT), a useful framework for understanding how people learn (Sweller, van Merriënboer, & Paas, 1998; van Merriënboer & Sweller, 2005), guided our design decisions during the development of ATTAC. CLT contends that as individuals gain expertise in a particular area or skill, they develop schemas that allow them to organize information in meaningful chunks. As expertise grows, they have more cognitive resources "freed up" to devote to challenging aspects of a task.

CLT posits that working memory capacity is limited and instructional systems should be designed to work within these constraints. According to CLT, there are three different types of cognitive load: extraneous, intrinsic, and germane load. Extraneous load arises from poor instructional design, such as a poor interface or including unnecessary information during the training. Intrinsic load stems from the instructional content itself, which is the number of "elements" learners must hold in their working memory at one time. The intrinsic load experienced by learners will vary based on the robustness of their schemas and is therefore highly dependent upon learners' prior knowledge. A more robust schema facilitates processing complex information more efficiently. Germane load refers to the level of mental effort learners must expend in order to make sense of the material they are learning and relate it back to their schema. These three types of cognitive load are traditionally thought of as additive, so once learners reach their individual capacity (i.e., cognitive overload), learning and performance suffers. Therefore, effective instruction should regulate the intrinsic load experienced by a student, minimize extraneous load, and promote germane load. Considering these principles, ATTAC adapts scenario difficulty and feedback based on user performance to achieve an efficient balance of cognitive load to maximize learning gains.

With CLT in mind, we designed our scenario difficulty adaptations to manage learners' cognitive load based on their performance during training. ATTAC includes scenarios with three difficulty levels: basic, intermediate, and advanced. Scenario difficulty is based on the number of elements the trainee needs to consider within the scenario (e.g., taking away JTAC tools, adding weather components, increasing the number of ordnance from which to select, etc.), such that more elements to consider yields more difficult scenarios. In addition, we also asked SMEs to rate the difficulty of our scenarios, and our difficulty ratings closely mapped to those of the SMEs. All trainees begin training at an intermediate difficulty. After two scenarios, their performance is assessed and the scenario difficulty is adjusted either up if trainees are performing well, down if trainees are performing poorly, or remains at the same level if trainees' performance is mixed. By adjusting scenario difficulty, ATTAC attempts to manage the intrinsic load for less knowledgeable trainees, so that they do not become overwhelmed with scenarios that are too complex and beyond their level of ability. Likewise, ATTAC provides more knowledgeable trainees with more complicated scenarios to prevent them from becoming bored.

Another finding in the CLT literature, the expertise reversal effect (ERE), guided the design of the feedback adaptation in ATTAC. The ERE suggests that certain instructional interventions that are effective for beginners may be detrimental for advanced students (Kalyuga, Ayres, Chandler, & Sweller, 2003). For example, Kalyuga, Chandler, and Sweller (2001) found that providing more detailed instruction was beneficial for novice learners, by giving them useful information about the task, but as learners gained more expertise about the domain, providing the detailed instruction hurt their performance. In fact, providing less structured instruction was beneficial for the experts. Kalyuga, Chandler, and Sweller (2001) argue that the detailed instruction led to extraneous cognitive load for the experts, as the additional detail was unnecessary and distracting for them, using up limited cognitive resources that could have otherwise been devoted toward meaningful cognitive processing. The ERE demonstrates how tailoring instruction to the needs of an individual learner is beneficial, as it is important to consider a learner's prior knowledge of the domain as some strategies may be more or less effective depending upon his or her level of expertise.

Therefore, ATTAC adapts feedback by adjusting the type of feedback trainees receive based on their scenario performance. For ideal game plans that meet Commander's intent, ATTAC delivers minimal outcome feedback indicating the answer was correct (i.e., "Good job!"). For acceptable game plans, the trainee is presented with outcome feedback (i.e., "Almost there. Your game plan would work but consider the following adjustments"), an ideal game plan, an error specific explanation of the flaws in the given game plan, and a summary of why the ideal game plan was optimal. For unacceptable answers, ATTAC provides outcome feedback (i.e., "Your game plan is unlikely to be successful"), an ideal game plan, and detailed process feedback that explains the thought process surrounding the optimal game plan. Adapting feedback in this manner helps address the ERE by tailoring how much feedback is being provided to trainees at different levels of performance. By limiting the amount of feedback provided to high performers, ATTAC reduces the amount of extraneous load they experience. Furthermore, by providing more feedback to lower performers, ATTAC promotes germane load by fostering schema development and helps these trainees understand why their answers were not ideal.

To test whether the AT techniques used in ATTAC led to better learning outcomes, we designed an experiment with three conditions: training with ATTAC, training with a non-adaptive version of ATTAC, and a minimal training control. Based on CLT and ERE, we hypothesized that an adaptive training approach would result in higher learning gains from pre- to post-test than a similar, non-adaptive approach and the minimal training control group.

METHOD

Design

Participants were randomly assigned to one of three between-subjects conditions. In the adaptive condition, participants received the full version of ATTAC which adapted feedback and difficulty based on an assessment of the student's game plan. In the non-adaptive condition, participants received a version of ATTAC that did not adapt either feedback or difficulty. Regardless of the game plan entered, participants' feedback included an ideal game plan to compare to the game plan they entered and the scenario difficulty remained the same throughout. In the control condition, participants received a short two-page refresher on game plan development, but did not perform any scenarios in ATTAC.

Participants

To date, 52 students enrolled in the Joint Fires Observer (JFO) course at the Expeditionary Warfare Center Training Group Pacific (EWTGPAC) participated in the experiment. All participants completed relevant lectures on game plan development from their course instructors prior to participating in the experiment. Of our sample population, 85% were in the USMC, and the remaining 15% were a mix of U.S. Navy, Air Force, and Army. All of the participants were male, their ages ranged from 19 to 38 years (M = 23.69, SD = 4.13), and their number of years in the military ranged from 1 to 15 years (M = 4.13, SD = 3.25). Approximately 73% of participants had a high school degree as their highest level of education, 10% had completed an associate's degree, 15% a bachelor's degree, and 2% a master's degree. Fifty-one percent of participants indicated they had no previous CAS experience, but 72% indicated they had previous call for fire experience, a related role that involves coordinating ground target attacks for artillery missions.

There were 20 students in the adaptive condition, 18 in the non-adaptive condition, and 14 in the control condition. Conditions did not significantly differ in terms of age, $F(2, 51) = 1.57$, $p = .22$, military experience, $F(2, 51) = 1.77$, $p = .18$, education level, χ^2 (6, $N = 52$) = 8.93, $p = .18$, subjective rating of CAS execution knowledge, $F(2, 51) = 0.67$, $p = .52$, previous CAS experience, χ^2 (2, $N = 52$) = 4.02, $p = .13$, or previous call for fire experience, χ^2 (4, $N = 52$) = 2.39, $p = .66$.

After preliminary data analysis, 8 participants were removed from the final data set. One participant was removed for failing to follow instructions on the pre and post-test; 3 participants were removed for having gain scores that were two standard deviations above and below the mean (e.g., did not finish post-test); 2 participants completed a number of training scenarios that were two standard deviations above the mean (i.e., over 80 scenarios in 35 minutes-indicating they were just clicking through the scenarios); and 2 participants were removed from the adaptive condition because they did not receive appropriate exposure to the treatment (i.e., they only adapted scenario difficulty once).

The final data set used for subsequent analyses included 44 participants. Fifteen individuals were in the adaptive condition, 16 in the non-adaptive condition, and 13 in the control condition. It was confirmed that the training conditions remained statistically equivalent on demographic measures after removal.

Materials

Demographic Questionnaire

The demographic questionnaire included items about participants' backgrounds (e.g., age, sex, education level), military experience (e.g., branch, occupational specialty, years in the military), and specific experience with call for fire and close air support. Participants also rated their agreement with the statement "I consider myself to be very knowledgeable about CAS execution" on a 1 (strongly disagree) to 5 (strongly agree) scale.

Pre- and Post-Tests

Two test forms were used for pre- and post-tests. Each test form included 9 scenarios similar to those used in ATTAC. Both forms contained identical items but were presented in a different order. The test forms used for pre- and post-test were counterbalanced such that each participant received one of the forms for the pre-test, and the other form for the post-test. To administer the tests, participants received a question packet with the scenarios and an answer packet, in which they circled their game plan responses for each question.

System Usability Scale (SUS)

The SUS is a 10-item survey that asks participants to indicate their agreement on a 1 (strongly disagree) to 5 (strongly agree) scale regarding the usability of the training environment (e.g., "I thought the system was easy to use;" Brooke, 1996). The SUS was administered to ensure that the ATTAC interface was intuitive and not a source of extraneous load.

Instruction Reactions Questionnaire (IRQ)

The IRQ is a 14-item survey developed for this study that asks participants to rate their agreement regarding their perceptions of the training on a 1 (strongly agree) to 6 (strongly disagree) scale. Example items include "Overall, the training was useful to me" and "The feedback I received was easy to understand." The IRQ was administered to gauge whether participants felt the training was helpful and appropriate.

Procedure

After obtaining consent for their participation in the research study, participants filled out the demographics questionnaire. Next, participants were given 15 minutes to complete the paper-based pre-test. After the pre-test, the procedure for the different conditions diverged, as depicted in Figure 2. Participants in the control condition received a two-slide refresher on game plan development before they moved onto the post-test. Participants in the adaptive and non-adaptive conditions interacted with laptops containing a tutorial about how to use ATTAC, a two-slide refresher on game plan development, and the condition-specific version of ATTAC. After going through the tutorial and refresher, participants in the adaptive and non-adaptive conditions trained using the system for 35 minutes. Once training was completed, participants in the adaptive and non-adaptive conditions filled out the SUS and IRQ. All participants received 10 minutes to complete the post-test, which contained the same 9 items from the pre-test but presented in a different order. Once the post-test was completed, participants in the control condition had the opportunity to interact with the adaptive version of ATTAC and completed the IRQ and SUS. Finally, all participants were debriefed and dismissed.

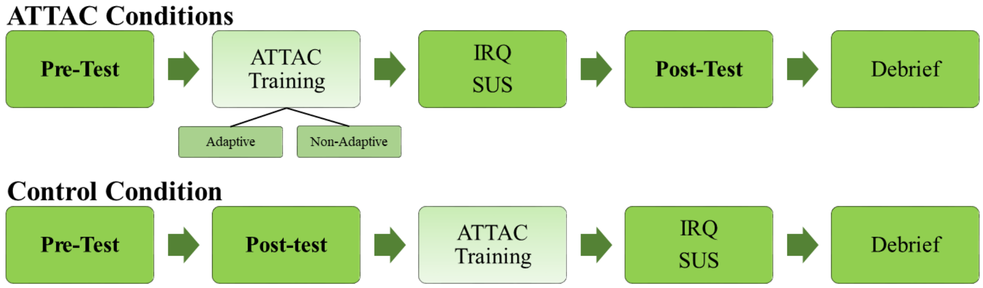

Figure 2. Experimental Procedure for the Three Conditions

RESULTS

Gain Score

Learning gain is a measure of how much trainees learned after taking into account the potential for what they could have learned. To this end, learning gain scores were computed for each trainee by calculating the difference between pre- and post-test scores and dividing by the difference between total possible score (126) and the pre-test score. Of practical significance when interpreting data, researchers often consider the effect size as a measure of the impact or magnitude of a particular result. In this case, we used Cohen's *d*, which is the size of the difference between two group means. According to Cohen (1988), a medium effect size is greater than 0.50 and a large effect size is greater than 0.80.

A one-way ANOVA was conducted using training condition as the between-subjects factor and the learning gain score as the dependent measure. As depicted in Figure 3, there was a significant main effect of condition, $F(2, 43) = 4.22$, $p = .02$, $\eta_p^2 = .17$. Post hoc comparisons revealed that those in the adaptive condition ($M = .35$, $SD = .19$) had a significantly higher learning gain score than those in the control condition with an effect size well exceeding Cohen's convention for a large effect ($M = .07$, $SD = .19$; $p = .008$, Cohen's $d = 1.46$). Likewise, those in the adaptive condition had a significantly higher learning gain score than the non-adaptive condition with a medium to large effect size ($M = .16$, $SD = .36$; $p = .048$, $d = 0.67$). The gain score in the non-adaptive group and the control group were not significantly different from one another ($p = .38$, $d = 0.30$), resulting in a small effect size. In summary, the results were consistent with our hypothesis that adaptive training would be more effective than non-adaptive training and the control.

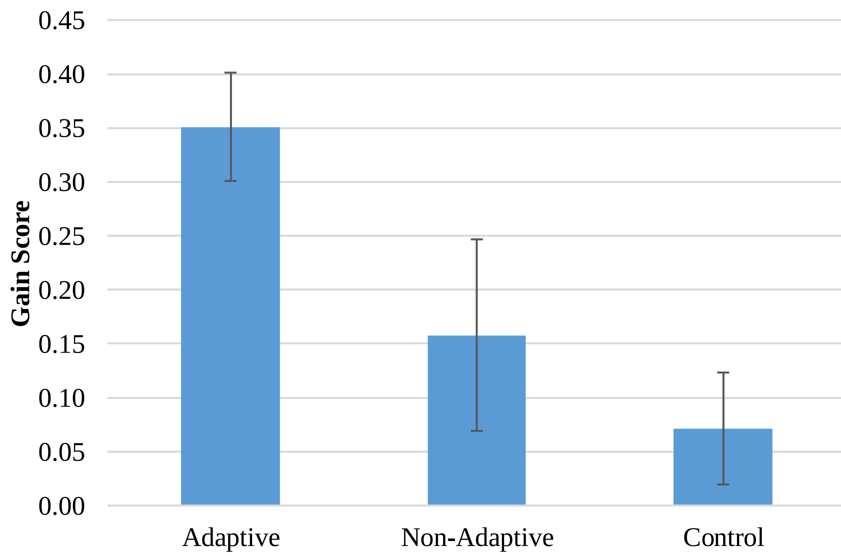

Figure 3. Learning Gains by Condition

SUS

Higher SUS scores indicate higher usability, and a SUS score of 68 is considered above average (Brooke, 1996, 2013). An independent samples *t*-test was conducted using training condition (i.e., the adaptive and non-adaptive conditions) as the grouping variable. Although the control condition also completed the SUS and IRQ after interacting with the adaptive version of ATTAC, their responses were not analyzed due to individuals in this condition interacting with ATTAC for different amounts of time. The overall usability ratings of the adaptive version ($M = 80.67$, $SD = 16.46$) and non-adaptive version ($M = 74.06$, $SD = 15.99$) of ATTAC were not significantly different $t(29) = 1.13$, $p = .27$ and both exceeded the above average SUS score. These results indicate that the ATTAC interface was designed such that the participants did not have any trouble understanding and interacting with the system.

IRQ

Lower IRQ ratings indicate higher level of agreement. For instance, scores below "3" indicate some level of agreement. Independent samples *t*-tests were conducted using training condition (i.e., the adaptive and non-adaptive conditions) as the grouping variable. Contrary to the findings for gain score, none of the 14 items on the IRQ differed significantly by condition. Five items are highlighted in Table 1, as they reflect the larger subset of questions. Overall, the adaptive and non-adaptive groups both had favorable attitudes toward the training, as both group means were below "3" across the three items meant to gauge participants' perception of the training experience. Additionally, groups did not differ regarding their beliefs that the scenarios were at an appropriate level of difficulty and that the feedback helped focus their attention on learning strategies to perform the task. Although not significant, those in the non-adaptive condition tended to have a more negative perception of the feedback than those in the adaptive condition.

Table 1. Summary of IRQ Means and SDs by Condition

Item	Adaptive Condition	Non-Adaptive Condition	*p*-value
"I liked the content in this training."	2.60 (1.84)	2.44 (1.37)	.78
"I would recommend this training to other trainees."	2.60 (1.81)	2.81 (1.47)	.72
"Overall, the training was useful to me."	2.29 (1.49)	2.44 (1.26)	.77
"The difficulty of scenarios was appropriate for my skill level."	2.80 (1.74)	3.00 (1.41)	.73
"I believe that the feedback I received focused my attention on learning strategies to perform this task better."	2.93 (1.71)	3.67 (1.59)	.23

Note: Descriptive statistics are displayed as *M* (*SD*).

DISCUSSION

The goal of this research was to investigate the efficacy of adapting feedback and scenario difficulty based on trainee performance to train a complex decision-making task. Applying the AT techniques used by Landsberg and colleagues (2012) to a new task, we developed a testbed called ATTAC for training game plan development (i.e., decision-making skills). Using principles from CLT and ERE to guide our design decisions, we developed ATTAC to manage the cognitive load of trainees to support more effective training, tailored to the individual trainee.

In an ongoing experiment with JFO students, preliminary data show that students who used the adaptive version of ATTAC had significantly higher learning gain scores from pre- to post-test than those who received either the non-adaptive version of ATTAC or those who received minimal additional training. These results are consistent with our hypothesis and the results of Landsberg et al. (2012) and CLT (Sweller, van Merriënboer, & Paas, 1998; van Merriënboer & Sweller, 2005). Adapting feedback and scenario difficulty based on a trainee's performance during training did result in more effective training than non-adaptive and traditional training approaches. After only 35 minutes of training, the adaptive version of ATTAC showed learning gains 400% higher than a minimal training control group, and 118% higher than a similar, but non-adaptive version of ATTAC.

By adapting feedback, ATTAC delivers feedback appropriate to the individual trainee's level of performance. When trainees perform poorly, the feedback is elaborative to enable the trainees to repair their understanding of the material. On the other hand, if trainees perform well, the feedback is minimal, presuming that the trainee understood the information well enough to derive the correct answer and avoid unnecessary cognitive load. Likewise, we adapted scenario difficulty to manage the intrinsic load of trainees. When a trainee is excelling, the scenario difficulty increases to provide challenge and prevent boredom. When a trainee is struggling, the scenario difficulty decreases to reduce intrinsic load and free up more cognitive resources for schema development and enabling the ability to tackle scenarios that are more challenging in the future. These preliminary data support the use of these types of AT techniques to train a complex decision-making task like game plan development.

Although there were significant differences in effect sizes across the three training groups, subjective ratings of ATTAC did not differ between the adaptive and non-adaptive conditions. Usability scores for both conditions were similar and relatively high, suggesting that the training interface did not induce undue extraneous load on the participants. With regard to the trainees' subjective reactions to the training, the pattern of results was not consistent with those for gain score. Participants did not rate the adaptive condition more favorably than the non-adaptive condition. In fact, both versions of ATTAC received generally favorable scores. A potential reason for this lack of differences is that both the adaptive and non-adaptive versions of ATTAC allow students to receive additional reps and sets to hone their game plan decision-making skills outside of the classroom. Many participants expressed a desire to use ATTAC after completing their course, including those in the non-adaptive condition.

Limitations and Lessons Learned

Overall, these initial findings suggest that implementing AT techniques may be effective for training decision-making skills. The goal of our experiment was to determine the effectiveness of particular AT techniques in a domain that has not been investigated previously. Other research investigating AT techniques have shown potential benefits (see Durlach & Ray, 2011), but suffered from poor comparison groups from which to make conclusions. Within the limited population from which we were able to recruit, we demonstrated that an AT approach was more effective than a similar, non-adaptive approach that provided repetition training only. However, the present study is limited because we were not able to explore systematically the unique contributions of the two adaptive techniques used in ATTAC. In future research, we plan to disentangle the effects of adaptive feedback and scenario difficulty have on learning in this domain.

Identifying effective adaptive techniques is critical for the future development of cost-effective training systems. A significant obstacle to developing and scaling AT solutions like ATTAC is the time associated with developing content. AT approaches require significant time and effort to ensure quality content. For ATTAC, each scenario required manually scoring each game plan to account for the many ways a JTAC could generate an effective game plan. Across the 90 scenarios in ATTAC, this resulted in flagging 1300 different game plan combinations as ideal or acceptable responses. In addition to the manual scoring process, feedback statements were individually written for these combinations so that trainees submitting acceptable and unacceptable responses received information that was relevant to the particular scenario they had just completed. In spite of these challenging development efforts, it is clear that adaptive approaches to training should have a place in the arsenal of learning tools available to instructors and students. However, these decisions must be made carefully, as there are considerable costs to develop effective AT systems. Therefore, it is essential that research continue to investigate how and when to best implement AT techniques and continue to identify the domains for which it is most effective in order to maximize learning gains and the effort required to develop quality training.

CONCLUSION

In this experiment, we demonstrated that AT techniques including adapting feedback and scenario difficulty were more effective than traditional one-size-fits-all approaches for training CAS game plan development, a complex decision-making task. When used as a supplement to traditional classroom instruction, AT approaches such as those used in ATTAC can help train students of all experience levels to a common baseline knowledge of the fundamentals of complex decision-making. This in turn allows instructors to focus on applications that are more complicated and increase the effectiveness of training. AT offers a more learner-centered approach by attempting to break the "brick and mortar" paradigm of traditional classroom instruction consistent with the U.S. Marine Corps (USMC) *Commandant's Planning Guidance.*

ACKNOWLEDGMENTS

We gratefully acknowledge Dr. Peter Squire and the Office of Naval Research who sponsored this work (Funding Doc# N0001418WX00298). We also wish to thank Dr. Shannon Bailey for her significant contributions during testbed development, and Mr. Jason Erbecker and Mr. Josh Foster for their subject matter expertise. Presentation of this material does not constitute or imply its endorsement, recommendation, or favoring by the U.S. Navy or Department of Defense (DoD). The opinions of the authors expressed herein do not necessarily state or reflect those of the U.S. Navy or DoD.

REFERENCES

Berger, D. H. (2019). Commandant's planning guidance. *United States Marine Corps.* Retrieved from https://www.hqmc.marines.mil/Portals/142/Docs/%2038th%20Commandant's%20Planning%20Guidance_2019.pdf?ver=2019-07-16-200152-700

Billings, D. R. (2012). Efficacy of adaptive feedback strategies in simulation-based training. *Military Psychology, 24*, 114–133.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, *13*(6), 4-16.

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability Evaluation in Industry*, *189*(194), 4-7.

Brooke, J. (2013). SUS: A retrospective. *Journal of Usability Studies, 8*(2), 29-40.

Conati, C., & Manske, M. (2009). Evaluating adaptive feedback in an educational computer game. In Z. Ruttkay, M. Kipp, A. Nijholt, & H.H. Vilhjalmsson (Eds.), *Intelligent virtual agents* (pp. 146-158). Berlin Heidelberg: Springer.

Cohen, J. (1988). *Statistical power analysis for the behavioral* sciences (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Durlach, P. J., & Ray, J. M. (2011). *Designing adaptive instructional environments: Insights from empirical evidence.* (Technical Report 1297). Arlington, VA: U.S. Army Research Institute for the Behavioral & Social Sciences.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23–31.

Kalyuga, S., Chandler, P., & Sweller, J. (2001). Learner experience and the efficiency of instructional guidance. *Educational Psychology*, *21*, 5-23.

Kelley, C. R. (1969). What is adaptive training? *Human Factors*, *11*(6), 547-556.

Landsberg, C. R., Mercado, A. D., Van Buskirk, W. L., Lineberry, M., & Steinhauser, N. (2012). Evaluation of an adaptive training system for submarine periscope operations. *Proceedings of the Human Factors and Ergonomics Society*, *56*(1), 2422-2426.

Landsberg, C. R., Van Buskirk, W. L., Astwood, R. S., Mercado, A. D., & Aakre, A. J. (2011) *Adaptive training considerations for simulation-based training systems* (Special report 2010-001). Defense Technical Information Center (DTIC).

McCarthy, J. E. (2008). Military applications of adaptive training technology. In M.D. Lytras, D. Gasevic, P. Ordonez de Pablos, & W. Huang (Eds.), *Technology enhanced learning: Best practices* (pp. 303-347). Hershey: IGI Publishing.

Serge, S. R., Priest, H. A., Durlach, P. J., & Johnson, C. I. (2013). The effects of static and adaptive performance feedback in game-based training. *Computers in Human Behavior*, *29*(3), 1150-1158.

Sweller, J., van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251-296.

van Merriënboer, J. J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(2), 147-177.