Generalizability of the Deliberate Practice Technique for Training Higher-Order Skills

Jeffrey M. Beaubien, Ph.D., Amy L. Alexander, Ph.D.
Aptima, Inc.
Woburn, MA
jbeaubien@aptima.com, aalexander@aptima.com

Sharnnia Artis, Ph.D.
Aptima, Inc.
Dayton, OH
sartis@aptima.com

Kara L. Orvis, Ph.D., Yale Marc, M.Arch.
Aptima, Inc.
Woburn, MA
korvis@aptima.com, ymarc@aptima.com

ABSTRACT

Previous research clearly demonstrates that expertise is directly related to the accumulated amount of deliberate practice, a training technique which allows knowledge to be encoded around key domain-related concepts and solution procedures, thereby facilitating rapid and reliable information retrieval (Ericsson, Krampfe & Tesch-Romer, 1993; Ericsson & Lehman, 1996). Until recently, there have been very few computer-based training programs that leverage the deliberate practice approach for training higher-order skills such as leadership, situation awareness, and decision-making. In this paper, the authors present interim training evaluation results from two such training efforts. These include *Wisdom: Lessons Learned from Operation Iraqi Freedom*™, which focuses on small unit Army leadership skills in Support, Stability, Transition, and Reconstruction (SSTR) operations, and *IMPACT: Team Skills Training for the Dynamic Targeting Cell*™, which focuses on Air Force leadership, teamwork, and decision-making skills in Dynamic Targeting Cell (DTC) operations. Quantitative and qualitative results suggest that both training programs produce practically and statistically significant gains in learning and self-efficacy. The results also suggest that the training programs are perceived as being practical and operationally relevant. Implications for low-fidelity simulations, in general, and deliberate practice-based simulations, in particular, are offered.

ABOUT THE AUTHORS

Jeffrey M. Beaubien is an Industrial and Organizational Psychologist at Aptima, Inc. For the past 11 years, his work has focused on developing new techniques for improving the effectiveness of simulation-based training. His work has been sponsored by the Federal Aviation Administration, the US Army, the US Air Force, and the TriCARE Management Authority, among others. Dr. Beaubien received a Ph.D. in Industrial and Organizational Psychology from George Mason University, an M.A. in Industrial and Organizational Psychology from the University of New Haven, and a B.A. in Psychology from the University of Rhode Island.

Amy L. Alexander is a Human Factors Scientist at Aptima, Inc. For the past 4 years, her work has focused on developing training programs and tools to support individual and team performance in the Combined Air and Space Operations Center (CAOC). Her work has been sponsored by the US Air Force, the US Navy, and the National Aeronautics and Space Administration, among others. Dr. Alexander received a Ph.D. in Engineering Psychology from the University of Illinois at Urbana-Champaign, an M.A. in

Engineering Psychology also from the University of Illinois, and a B.S. in Psychology from the Ohio State University.

Sharnnia Artis is a Human Factors Engineer at Aptima, Inc. with experience in instructional systems design, user-centered software design, and cultural ergonomics. Dr. Artis' work includes developing scenario-based training programs for critical cognitive skills, developing measures of individual and team performance in Air Operations Centers, and evaluating the effectiveness of blended learning programs. Dr. Artis holds a Ph.D. in Industrial and Systems Engineering from the Virginia Polytechnic Institute.

Kara L. Orvis is an Industrial and Organizational Psychologist at Aptima, Inc.. Dr. Orvis has authored over 60 publications and professional presentations in the areas of team leadership, dispersed team collaboration, dispersed leadership, and training technologies. Her work has been sponsored by the US Army, the US Air Force, and the US Navy, among others. Dr. Orvis holds a Ph.D. in Industrial-Organizational Psychology from George Mason University, an M.A. In Industrial-Organizational Psychology also from George Mason University, and a B.A. in Psychology from Ohio Wesleyan University.

Yale Marc is an Interactive Training Specialist at Aptima, Inc. For the past 18 years, his work has involved the design, development, and delivery of interactive training programs for government, business, and educational clients. His work has been sponsored by the US Air Force, the US Army, Cisco Systems, Motorola, Nokia, and Hewlett-Packard, among others. Mr. Marc received master's degrees from Tulane University and Harvard's Graduate School of Design. He is currently working on his doctorate in Educational Technology.

Generalizability of the Deliberate Practice Technique for Training Higher-Order Skills

Jeffrey M. Beaubien, Ph.D., Amy L. Alexander, Ph.D.
Aptima, Inc.
Woburn, MA
jbeaubien@aptima.com, aalexander@aptima.com

Sharnnia Artis, Ph.D.
Aptima, Inc.
Dayton, OH
sartis@aptima.com

Kara L. Orvis, Ph.D., Yale Marc, M.Arch.
Aptima, Inc.
Woburn, MA
korvis@aptima.com, ymarc@aptima.com

INTRODUCTION

Research from the field of naturalistic decision making clearly demonstrates that the cognitive processes of experts are qualitatively different than those of novices. When encountering new problems, experts tend to draw heavily on their prior experience rather than relying on "textbook" solutions. Additionally, a substantial proportion of the expert's time is spent assessing the situation and looking for deep underlying patterns within the problem space, rather than focusing solely on the problem's surface features. This decision-making process generally results in the expert selecting a very limited number of plausible courses of action and evaluating them, often in parallel, by means of mental simulation. Finally, experts frequently "satisfice" – selecting an effective approach rather than seeking an optimal solution (Klein, 1995, 1997).

Previous research also demonstrates that expertise is directly related to the accumulated amount of deliberate practice, a training technique which allows knowledge to be encoded around key domain-related concepts and solution procedures, thereby facilitating rapid and reliable information retrieval (Ericsson et al., 1993; Ericsson & Lehman, 1996). Deliberate practice differs from other training approaches in several critical ways. Specifically, deliberate practice: 1) involves a high degree of repetition to develop expert habits that are near-automatic in their execution; 2) involves focused feedback to help learners better target their areas of weakness, thereby conserving limited training resources; 3) involves a series of short, stop-and-start exercises to identify and resolve performance problems early; 4) emphasizes difficult, rather than mundane, situations for which the learner is likely to be unprepared and which will stimulate high levels of motivation, and; 5) involves active coaching with a high instructor-to-student ratio (Ericsson et al., 1993; Ericsson & Lehmann, 1996; Lussier, Shadrick & Prevou, 2003).

Until recently, there have been very few computer-based training programs that leverage the deliberate practice approach for training higher-order skills such as leadership, situation awareness, and decision making. One of these, *Red Cape: Crisis Action Planning and Execution*, was developed to train military-civilian inter-agency coordination during homeland security and natural disaster crises. A rigorous training evaluation clearly demonstrated that *Red Cape* trainees – including a mix of Indiana Army National Guard and Indiana Department of Homeland Security personnel – liked the training, found it to be useful and operationally relevant, and demonstrated both practically and statistically significant gains in learning (Beaubien, Shadrick, Paley, Badugu, Ennis, & Jacklin, 2006; Schaefer, Shadrick, Beaubien, & Crabb, 2008; Shadrick, Schaefer, & Beaubien, 2007).

Although the initial training evaluation results were impressive, the research team was unsure about the extent to which the observed findings would generalize beyond the original sample and with new training content. Therefore, additional follow-up training evaluation studies were clearly required.

At the same time, our personal observations and experiences revealed that the initial *Red Cape* training software had several critical shortcomings that limited its effectiveness and ease of re-use. First, the user interface appeared somewhat dated by modern standards. This lowered the training's perceived face validity and otherwise detracted from the quality of the learning modules and expert feedback. Second, the software provided no user login screen. This made it difficult for the learners to pause their work and resume at a later date, which is necessary to capitalize on the effects of distributed practice.

Third, the training software relied heavily on a physically-present instructor to help impart key lessons learned at the end of each learning module. As a result, the software did not support self-directed or instructorless learning. Fourth, the training software did not provide cumulative feedback to help the user better understand his or her unique profile of strengths and weaknesses. As a result, learners sometimes placed greater weight on the performance feedback that they had received from their most recent learning module, even if it was not necessarily reflective of their cumulative performance. Finally, the training software was not SCORM-conformant. As a result, the learning modules could not be re-used or incorporated into Department of Defense (DoD) learning management systems.

With these limitations in mind, the research team completed a major overhaul of the software application. First, the software was completely re-written using Adobe FLEX, which is specifically designed for multimedia applications. Doing so allowed the researchers to incorporate advanced features, such as embedded multimedia controls and high-resolution video. The end result is a more modern, visually appealing, and streamlined user experience. Furthermore, the new software application allowed the researchers to develop a login feature to track the learner's progress and provide the learner the opportunity to pause and resume their work during multiple sittings. Second, expert feedback was embedded within the training software, rather than providing it in a separate instructor guide. As a result, the software now supports both instructor and instructorless training modes, should an instructor/facilitator not be available.

Third, the researchers developed a combination of scenario-specific and cumulative feedback screens that provided the learners a comprehensive depiction of their performance to ensure a highly efficient training experience. Specifically, the software now includes focused feedback on their strong points and areas of weakness on individual scenarios and cumulative feedback that encompass the learner' performance across all of the completed learning modules completed to date. Fifth and finally, the authors made the training software fully SCORM-conformant, thereby permitting the training content to be re-used in any DoD learning management system.

The training software operates as follows: First, the learner is presented with an introduction module that provides factual knowledge about the critical skills that are to be trained – such as leadership and communication – and their relevance to the current operational environment. Whenever possible, this information is narrated by recently-retired Subject Matter Experts (SMEs) to make the training content more engaging and face valid (see Figure 1).

Figure 1. IMPACT^{TM} – Introduction Module

Next, learners complete a series of learning modules which require them to apply the critical knowledge that they have just learned. Each learning module begins with a professionally-narrated, 3-5 minute FLASH video of a realistic tactical problem. Battlefield maps and other visual elements, such as JADOCS coordination screens, are used to provide additional realism and to further the storyline (see Figure 2).

Figure 2. IMPACT™ – Learning Module

After watching the video, learners use the software's built-in notes screen to describe how they would apply the critical skills in that particular situation. Learners are provided with a number of tools to assist them in their decision-making process, such as the definition of each critical skill, a Commander's mission intent statement, decision support matrices, as well as the narration script. Learners are also prompted to respond in a timely manner with the aid of a built-in timer that appears in the center of the screen (see Figure 3).

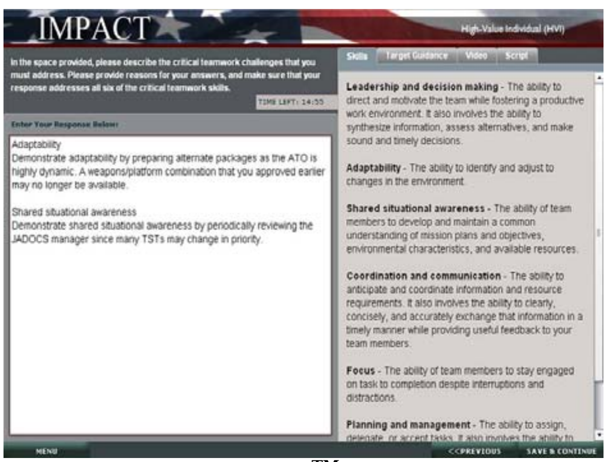

Figure 3. IMPACT™ – Notes Screen

After saving their responses, learners are presented with a series of behaviors that our SMEs would have performed in that situation. Learners are instructed to check off each behavior that is reflected in their notes, thereby self-assessing their performance. While not explicitly represented to the learner, all behaviors are mapped to the higher-order skills being trained (see Figure 4). Some of these behaviors are "red herrings" that prevent learners from "gaming the system" because simply checking off all behaviors will result in a less-than-optimal outcome.

Figure 4. IMPACT™ – Self-Assessment Screen

Once learners have saved their self-assessments, the software identifies those skills which they did well on (colored in green), as well as those skills which require additional attention (colored in yellow and red, respectively). Learners must select at least 76% of the relevant behaviors to be rated as "excellent," 51-75% to be rated as "good," and 50% or less to receive a "needs work" rating (see Figure 5).

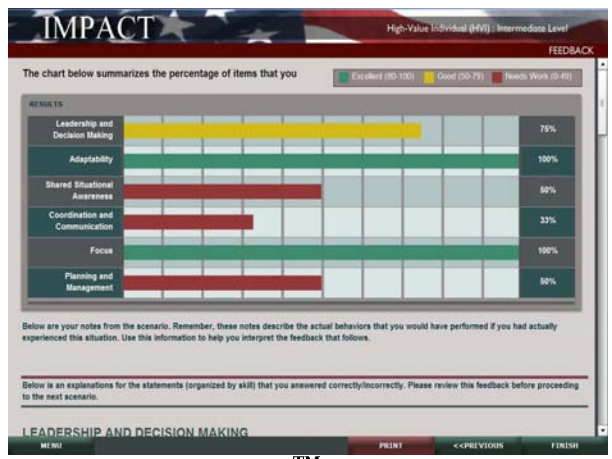

Figure 5. IMPACT™ – Feedback Screen

Directly below the graphical feedback, the software provides diagnostic feedback to help the learners better understand why those behaviors were correct or incorrect within the previous scenario. Correct responses are preceded by a green "✓" and incorrect responses are preceded by a red "×" (see Figure 6). The feedback can be printed out and/or accessed by the learner at a later time, should the learner want to do so.

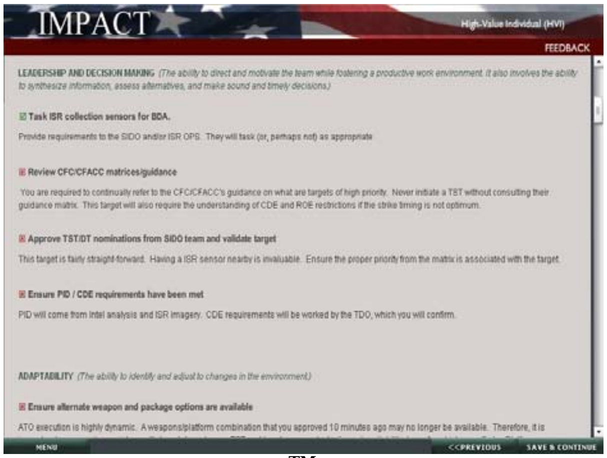

Figure 6. IMPACT™ – Feedback Screen

In the sections that follow, we present interim training evaluation results from two modifications of the initial *Red Cape* training software. The first training program, titled *Wisdom: Lessons Learned from Operation Iraqi Freedom™*, focuses on small unit Army leadership skills in Support, Stability, Transition, and Reconstruction (SSTR) operations. The second training program, titled *IMPACT: Team Skills Training for the Dynamic Targeting Cell™*, focuses on Air Force leadership, teamwork, and decision-making skills in Dynamic Targeting Cell (DTC) operations.

STUDY #1 METHOD

Participants

The participants included 97 NCOs and officers, including: 16 Sergeants, 39 Staff Sergeants, 24 Sergeants First Class, 2 Lieutenants, and 14 Captains. The mean age was 31.7 (SD = 5.2) years. The mean number of years military service was 10.5 (SD = 4.1), and the mean level of education was 14.0 years (SD = 1.5). The sample was overwhelmingly male (75.3%). Forty-four of the participants were from recruited from the Signal Corps (MOS 25). An additional 50 participants were recruited from Human Resources (MOS 42). Three additional participants did not list their Military Occupational Specialty (MOS). Unfortunately, none of the intended audience, Airborne Infantry (MOS 11), were available to participate.

Because the learning modules focused on small unit leadership in dismounted urban combat operations, there was a clear mismatch between the intended target audience and the participant sample for this training evaluation. That being said, many of the participants should have been familiar with the critical leadership skills that were being addressed – judgment, emotional maturity, cultural awareness, and self-control – because these skills were drawn directly from *Field Manual 22-100* (Department of the Army, 1999) which focuses on small unit leadership. Indirect evidence clearly supports our belief that they were familiar with the content, as an overwhelming majority of the participants were enrolled in the Army's Basic (n = 42) and Advanced NCO Courses (n = 39). In addition, over two-thirds of the participants (70.2%) had previously been deployed in support Operations Iraqi Freedom (OIF) or Enduring Freedom (OEF). Therefore, the participants should have been familiar with the learning content, even though they would not be expected to have formal experience in applying these skills to urban combat situations.

Materials

The research materials included a paper-and-pencil questionnaire and the *Wisdom™* computer-based training program. The questionnaire included items that inquire about basic demographic variables, along with several custom-developed items that inquire about the participants' task-specific self-efficacy (Bandura, 1982), and their perceptions of the *Wisdom™* training program's usefulness (Alliger & Janak, 1989; Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997). For ease of administration, the demographic questions were completed first. Next, the participants completed the *Wisdom™* training program. Finally, the participants completed the self-efficacy and perceived utility questions. Finally, they were given the opportunity to provide open-ended comments about the training program.

Measures

Self-efficacy is defined as the belief that one can successfully perform a series of specific behaviors that are relevant to success in one's job (Bandura, 1982). Self-efficacy influences the job-related behaviors that one chooses to engage in, the amount of effort which one exerts at work, and one's level of persistence in the face of work-related obstacles. Task-specific self-efficacy was measured with six positively-worded items that were specifically written for this study. Example items include *I am confident that I can perform the leadership skills that were taught in this training program* and *Having completed this training program, I'm certain that I can handle many different leadership challenges*. Each item was assessed using a 5-point Likert scale, with anchors that ranged from low (1) to high (5). The internal consistency reliability for this measure was $\alpha = .86$.

Perceived usefulness or utility is defined as the belief that the training was well-designed and job-relevant. Previous research (Alliger & Janak, 1989; Alliger et al., 1993) suggests that perceived usefulness is positively, albeit weakly, related to measures of learning. Perceived usefulness was measured with ten positively-worded items that were specifically written for this study. Example items include *The training helped me to meet my learning objectives* and *The leadership skills that I learned are the same as those that I expect to use in the field*. Each item was assessed using a 5-point Likert-type scale, with anchors that ranged from low (1) to high (5). The internal consistency reliability for this measure was α = .91.

Learning was defined as the percentage of expert behaviors that the participant answered correctly per learning module. Because there were six learning modules, there were six corresponding measures of learning per participant. The mean percentage items correct was 40.2 (SD = 19.3), 57.6 (SD = 20.4), 57.2 (SD = 22.1), 68.3 (SD = 19.3), 63.0 (SD = 21.9), and 67.0 (SD = 22.4) for learning modules 1 through 6, respectively.

Design and Procedure

Each participant was seated in front of a PC workstation that was equipped with a 19" LCD monitor and stereo headphones. The participants were given a short briefing about the research project, and were instructed on how to operate the training software. All participants then completed pages 1-4 of the paper-and-pencil questionnaire which inquired about their professional background, training, and experience. Next, the participants completed modules 1-6 from the *Wisdom™* training program in sequential order. Finally, each participant completed pages 5-8 of the paper-and-pencil questionnaire which contained randomly-ordered items about their post-training self-efficacy and perceived usefulness of the *Wisdom™* training tool. The entire testing process took approximately three hours.

STUDY #1 RESULTS

Learning

Measures of learning were assessed using repeated measures analysis of variance to determine if there was a statistically-significant trend across the 6 training modules. The data indicated a significant multivariate effect $F(5,88) = 20.26$, $p = .00$. Drill-down analyses indicated a significant quadratic effect $F(1) = 28.66$, $p = .00$, with performance improving in a near-linear fashion across learning modules 1-4, and then reaching a plateau. In essence, the participants started off at a mean level of 40% correct items, peaked at 68% correct items, and then leveled off. The results are depicted graphically in Figure 7 below.

Figure 7. Wisdom™ – Results

Self-Efficacy

Measures of task-specific self-efficacy were assessed using a 1-sample *t*-test that compared the self-efficacy scale's observed mean value of 3.82 (SD = .68) to a benchmark value of 3.0 which represents *Neither Agree nor Disagree* on the 5-point Likert scale. The effect was statistically significant, $t(91) = 11.5$, $p = .00$. The results indicated that participants believed that they could successfully apply their newly-trained skills to the post-training work environment.

Perceived Usefulness

Measures of perceived usefulness were assessed using a 1-sample *t*-test that compared the perceived utility scale's observed mean value of 3.50 (SD = .76) to a benchmark value of 3 which represents *Neither Agree nor Disagree* on the 5-point Likert scale. The effect was statistically significant, $t(91) = 6.3$, $p = .00$. The results indicated that participants found the *Wisdom™* training program to be well-designed, that it met their expectations, and that it was mission relevant.

Open-Ended Responses

In the post-training questionnaire, participants were provided the opportunity to more fully articulate their perceptions of the *Wisdom™* training tool. Much of feedback was positive. Below are a series of representative comments:

- The training made me realize that being culturally aware and maintaining self-control is vital to the completion and success of a mission. I know what I need to work on.
- It placed people in difficult situations. It didn't feel like a textbook.
- [The] scenarios were very realistic and hard hitting.
- It made me look at the bigger picture and options I wouldn't have even thought about.
- I liked the post-assessment feedback and other considerations. A lot of the information is information that I haven't considered [before].
- The after action tells you what you could have done better. It pointed out some mistakes and it briefly explained what you could have done better.
- The scenarios were pretty in-depth and they didn't just give you a situation to figure out on your own. They actually gave you the supporting facts.

When negative feedback was provided, it often addressed technical glitches with the software which were due to problems with some of the PC workstations in the electronic learning laboratory. Other negative comments reflected a fundamental disconnect between the training content and the learners' military Occupational Specialty (MOS), which was unfortunate but unavoidable. Below are a series of representative comments:

- Not germane to Signal MOS.
- I did not have any experience to pull from in my decision-making.
- I think that the 10 minute [limitation] was too short. I did not have time to write [my responses] and pick all the data.
- Even though the situations were relevant and up to date, a Soldier's actions in the situations are often times more than what a Soldier sitting at a computer would think to type in.

STUDY #1 CONCLUSIONS

Despite these limitations, the training evaluation results clearly demonstrate that the *Wisdom*™ training package produces both practically and statistically significant gains in learning. The results also demonstrate the *Wisdom*™ training tool improves the learners' self-confidence in their abilities, which is an important learning outcome in its own right (Kraiger, Ford, & Salas, 1993). Finally, the results demonstrate the learners perceive the *Wisdom*™ training tool as being relevant to the Army's ongoing missions in Iraq, Afghanistan and elsewhere.

These positive results are not entirely surprising, given that *Wisdom*™ leveraged the deliberate practice technique that has been used with great success on the Army's previously-validated *Red Cape* and *Think Like a Commander* training platforms (Beaubien, Shadrick, Paley et al., 2006; Lussier et al., 2003; Schaefer, Shadrick, Beaubien, Crabb, 2008; Shadrick & Lussier, 2004). What was surprising, however, was that we received such positive feedback from participants in the Signals Corps and Human Resources, because the learning modules were targeted to dismounted infantry units. We believe that the high levels of learning were due, in part, to the fact that the training content was highly realistic and that the leadership skills trained – including judgment, emotional maturity, cultural awareness, and self-control – were already familiar to the participants from their other NCO and officer training courses.

STUDY #2 METHOD

Participants

Training evaluation data were collected from 11 recently-retired Dynamic Targeting Cell (DTC) SMEs, none of whom had participated in the development of *IMPACT*™. The sample included 6 former DTC Chiefs or Deputy Chiefs, 4 Target Duty Officers, and 1 Attack Coordinator. Several officer ranks were represented in the sample, including 2 Air Force Colonels, 4 Lieutenant Colonels, 2 Majors, 1 Captain, and 2 Lieutenants. The mean number of active duty years of service was 22.6 (SD = 6.5). Most of the participants had completed either a Bachelor's degree (n = 4) or a Master's degree (n = 6). The sample was overwhelmingly male (90.9%).

Materials

The materials included a paper-and-pencil questionnaire and the computer-based training program titled *IMPACT: Team Skills Training for the Dynamic Targeting Cell*™. The questionnaire – which was nearly identical to that used in Study #1 – included items that inquire about basic demographic variables, along with several custom-developed items that inquire about the participants' task-specific self-efficacy (Bandura, 1982), and their perceptions of the *IMPACT*™ training program's usefulness (Alliger & Janak, 1989; Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997). For ease of administration, participants completed the demographic questions first. They then completed the *IMPACT*™ training program.

Finally, they completed the task-specific self-efficacy and perceived usefulness/utility questions.

Design and Procedure

The research design and procedure was very similar to Study #1, with a few notable differences. Perhaps the biggest difference is that there was a substantially larger number (n = 15) of learning modules for the participants to complete. The learning modules were organized into the 5 broad classes of DTC operations: theater ballistic missile launch, key command and control node, high-value individual, Commander's high-priority target, and weapons of mass effect. Each class of DTC operations included 1 "low difficulty," 1 "moderate difficulty," and 1 "high difficulty" learning module. Difficulty was systematically manipulated by altering key mission parameters such as environmental and weather conditions, target visibility, available weapon-target pairing, de-confliction requirements, and the like. Difficulty ratings were developed by consensus of 5 recently-retired DTC SMEs – including 1 Colonel, 2 Lieutenant Colonels, and 2 Majors – who designed the training content.

Each participant was seated in front of a PC laptop that was equipped with a widescreen LCD display and stereo headphones. The participants were given a short briefing about the research project, and were instructed on how to operate the *IMPACT*™ software program's user interface. All participants then completed a short questionnaire which inquired about their professional background, training, and experience.

Next, the participants reviewed the critical teamwork skills that were taught in the *IMPACT*™ training program. These materials were drawn directly from the Combat Operations Division's Mission Essential Competencies (MECs℠) effort, and therefore should have already been familiar to them. Next, the participants completed the first 7 learning modules (all 5 of the "low difficulty" modules and 2 of the "moderate difficulty" modules) during an 8-hour session. The learners then returned one week later to complete the remaining 8 learning modules (the 3 remaining "moderate difficulty" modules and all 5 of the "high difficulty" modules) during a separate 8-hour session. Finally, each participant completed a short questionnaire which contained randomly-ordered items about their post-training self-efficacy and perceived utility of the *IMPACT*™ training tool. The entire testing process took approximately 16 hours.

STUDY #2 RESULTS

As in the first study, 3 criterion measures were assessed: learning, perceived usefulness, and self-efficacy. Measures of learning – which was assessed as the percentage of expert responses correctly scored per learning module – were taken during each of the 15 learning modules. Measures of utility and self-efficacy – both of which were assessed on a 5-point Likert scale – were taken at the end of each data collection episode, and were spaced one week apart.

Learning

As before, learning was assessed using a repeated measures analysis of variance. Due to the small sample size (n = 11), three composite learning scores were created. The first measure was the mean percentage of items correct for the 5 "low difficulty" scenarios. The second and third measures were calculated as the mean percentage of items correct for the 5 "moderate difficulty" and 5 "high difficultly" scenarios, respectively. The mean percent correct for the "low difficulty," "moderate difficulty," and "high difficulty" learning modules was 81.0 (SD = 6.09), 91.3 (SD = 4.98); and 91.3 (SD = 6.15), respectively. A significant quadratic effect was detected ($F = 52.04$, $p < .01$), suggesting that the learners exhibited significant initial gains in learning from the low to medium difficulty scenarios, and then a plateau between the medium and high difficulty scenarios. The results are depicted graphically in Figure 7 below.

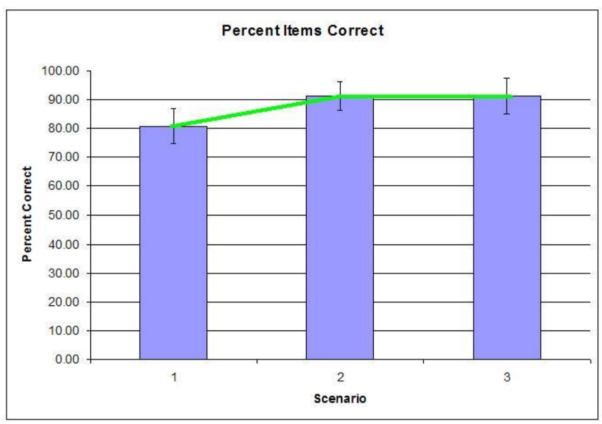

Figure 8. IMPACT™ – Results

Self-Efficacy

Measures of task-specific self-efficacy were again assessed using a 1-sample t-test, with a benchmark value of 3.0. At the end of the first data collection session, participants felt confident that they could apply the *IMPACT*™ trained skills in the post-training environment (M = 4.17, SD = .59, t = 6.8, $p < .01$). At the end of the second data collection episode one week later, participants again indicated that they felt confident in applying the *IMPACT*™ trained skills (M = 4.01, SD = .75, t = 4.4, $p < .01$). The difference between these values was not significant (t = 2.1, ns), indicating that the participants' initially high levels of self-efficacy did not significantly decrease over time.

Perceived Usefulness

Measures of perceived usefulness were also assessed using a 1-sample t-test, with a benchmark value of 3.0. At the end of the first data collection session, participants indicated that the *IMPACT*™ training was highly useful (M = 4.16, SD = .49, t = 8.2, $p < .01$). At the end of the second data collection session, participants again indicated that the *IMPACT*™ training was highly useful (M = 4.02, SD = .53, t = 6.4, $p < .01$). The difference between these values was not significant (t = 1.5, ns), also indicating that the participants' initially favorable responses did not significantly decrease over time.

Open-Ended Responses

In the post-training questionnaire, participants were provided the opportunity to more fully articulate their perceptions of *IMPACT*™. Much of feedback was positive, and many of the comments focused on the training's level of perceived realism and how the multimedia effects kept them engaged throughout the experience. Below are a series of representative comments:

- Just its existence is a good thing since there are few efforts out there that go to this level of specificity and detail for DTC.
- Good scenarios, realistic targets and dynamic situations. I liked the decision points (choices) and feedback on my choices.
- Scenarios were complete, with much specific information that directly impacted one's decisions. Good variety of scenarios.
- Thought-provoking scenarios. Liked how you need to analyze your thought process. Excellent graphics and good video clips. Nice amount of time, like that it is limited.
- Visual effects keep the training interesting and lively.

When negative feedback was provided, it often addressed the desire for additional functionality within the training software. Below are a series of representative comments:

- Include a canned scenario where students see step-by-step a successful as well as unsuccessful example. For example, show more screens of JADOCS and chat to show examples of both good and bad steps in the DTC process.
- Include a sample DTC demo in which your panel of experts walks through a vignette. Amplify [this] with panel members' thoughts, coordination, and information gleaning.
- Include more scenario-specific consideration in the survey stimulating deeper thought.
- Include a user "take-away" that lists generic things to consider during dynamic targets (e.g., weather, CDE, munitions, time constraints, guidance, threat, coordination with JTF counterparts, and dozens more).

STUDY #2 CONCLUSIONS

Despite the relatively small sample size, the training evaluation results clearly demonstrated that *IMPACT*™ produces both practically and statistically significant gains in learning. Although the learning effect did reach a plateau by the time learners had completed the "moderate difficulty" learning modules, this may represent a ceiling effect which was caused by the presence of very senior DTC personnel (who now serve as DTC trainers) in the participant sample. It is unclear to what extent these results would be replicated with more junior personnel. Additional data collection opportunities are currently being planned with active duty personnel, which will provide a much larger sample size and which will permit more detailed multivariate analyses. When complete, the two sets of results will be compared and contrasted.

Finally, the data suggest that that *IMPACT*™ training produced high levels of self-efficacy and was perceived as useful by the participants. Moreover, these positive findings did not deteriorate between the two data collection sessions. The authors believe that these positive findings are due, in part, to the highly realistic and engaging training scenarios that were developed by our in-house DTC SMEs.

GENERAL CONCLUSIONS

The findings presented here further support the efficacy of the deliberate practice approach to complex skill acquisition using low-fidelity, scenario-based training methods. To our knowledge, there have been four such efforts conducted to date, all of which have some degree of evidence supporting their effectiveness: *Think Like a Commander* (Lussier et al., 2003; Shadrick & Lussier, 2004), *Red Cape: Crisis Action Planning and Execution* (Beaubien, Shadrick, Paley, et al., 2006; Schaeffer, Shadrick, Beaubien, & Crabb, 2008), *Wisdom: Lessons Learned from Operation Iraqi Freedom™*, and *IMPACT: Team Skills Training for the Dynamic Targeting Cell™*. To date, however, there have been no long-term validation studies, thereby preventing the authors from making any firm conclusions regarding the long-term effectiveness of this approach.

Over the past several years, our research team has made numerous improvements to both the training methodology and the associated software tool because, at least in our view, these two components are highly-interrelated. Many of these changes – including the addition of increasingly more complex learning modules, post-scenario and cumulative feedback graphs, embedded diagnostic feedback for each critical behavior, the use of "red herrings" to prevent possible malingering, and full SCORM compliance – have been critical to enhancing the overall training experience, and better aligning the training methodology with the spirit of the deliberate practice approach.

That being said, there is still room for improvement going forward. Future versions of the training tool will likely include new functionality that will visually compare and contrast the learners' cumulative feedback with that from their most recent learning module. This is, in our view, one of the most critical issues to address going forward. Obviously, the deliberate practice approach relies on cumulative feedback to help the learner better understand his or her unique profile of strengths and weaknesses. However, one's cumulative feedback may conflict with the feedback from the most recent learning module. Therefore, new techniques for visually comparing and contrasting these two types of feedback are clearly required.

Another issue that needs to be addressed going forward is that of learner motivation. For all of its merits, the deliberate practice approach can be described as "negative" in that it always seems to focus on the learner's primary areas of weakness. This is purposely done to conserve limited training resources. However, a compelling body of research suggests that focusing on weaknesses may not be the best approach. Rather, attempting to instill "learning goals" – which encourage learners to explore the limits of their knowledge, and which view mistakes as a natural part of the learning process – may promote more effective long-term retention (Payne, Youngcourt, & Beaubien, 2007). To our knowledge, no attempts have yet been made to integrate learning goals in the deliberate practice paradigm.

In summary, the available evidence suggests that the deliberate practice approach is a promising technique for training higher-order skills using low-cost simulation-based methods. That being said, the researchers believe that enhancements can be made to this training approach and intend to address them in future training efforts going forward.

ACKNOWLEDGEMENTS

Wisdom: Lessons Learned from Operation Iraqi Freedom™ was developed by Aptima, Inc. under contract for the US Army Research Institute, Leader Development Research Unit. *IMPACT: Team Skills Training for the Dynamic Targeting Cell™* was developed by Aptima, Inc. under contract for the US Air Force Research Laboratory, Human Effectiveness Directorate. The views, opinions, and findings contained in this report are those of the authors and shall not be construed as an official Department of Defense position, policy, or decision, unless so designated by other documents.

REFERENCES

Alliger, G., & Janak, E. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology, 42*, 331-342.

Alliger, G., Tannenbaum, S., Bennett, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology, 50*, 341-358.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*, 122-147.

Beaubien, J., Shadrick, S., Paley, M., Badugu, S., Ennis, C., & Jacklin, S. (2006). Using deliberate practice to train military-civilian interagency coordination. In *Proceedings of the 2006 Interservice/Industry Training, Simulation, & Education Conference (I/ITSEC)*. Paper No. 2785. I/ITSEC: Orlando, FL.

Department of the Army. (1999, August). *Army leadership: Be, know, do*. Field Manual 22-100. Washington, DC: Author.

Ericsson, K., Krampe, R., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, 363-406.

Ericsson, K., & Lehmann, C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*, 273-305.

Klein, G. (1997). The recognition-primed decision (RPD) model: Looking back, looking forward. In C. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 285-292). Mahwah, NJ: Erlbaum.

Klein, G. (1995). A recognition-primed decision (RPD) model of rapid decision making. In G. Klein, J. Orasanu, R. Calderwood, & C. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 138-146). Norwood, NJ: Ablex.

Kraiger, K., Ford, J., & Salas, E. (1993). Application of cognitive, skill-based and affective theories of learning to new methods of training evaluation. *Journal of Applied Psychology, 78*, 311-328.

Lussier, J., Shadrick, S., & Prevou, M. (2003). *Think Like a Commander Prototype: Instructor's Guide to Adaptive Thinking* (Research Product No. 2003-2). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Payne, S.C., Youngcourt, S.S., & Beaubien, J.M. (2007). A meta-analytic examination of the goal orientation nomological net. *Journal of Applied Psychology, 92*, 128-150.

Schaefer, P., Shadrick, S., Beaubien, J., & Crabb, B. (2008). *Training Effectiveness Assessment of Red Cape: Crisis Action Planning and Execution*. Research Report No. 1885. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Shadrick, S., & Lussier, J. (2004). *Assessment of the Think Like a Commander Training Program* Research Report No. 1824. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Shadrick, S., Schaefer, P., & Beaubien, J. (2007). *Development and content validation of crisis response training package 'Red Cape: Crisis Action Planning and Execution.'* Research Report No. 1875. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.