Leveraging Machine Learning and Cognitive Science to Enhance Knowledge Retention in Air Force Special Warfare Trainees

Amy M. Smith, PhD
Blank Slate Technologies
New Haven, CT
amy@blankslatetech.co

ABSTRACT

At the core of every good decision is a strong knowledge base. Making fast and accurate decisions in high-stakes scenarios requires effortless and automatic recall of important information. Despite this fundamental truth, military training dedicated to cognitive human performance is underemphasized in training relative to physical human performance. To address this discrepancy in their medic training, the 351st Special Warfare Training Squadron (SWTS) at Kirtland Air Force Base explored the application of innovative learning techniques from cognitive science with their pararescue trainees in 2022. Over the course of one year, 76 trainees across three classes at the 351 SWTS used a mobile application that combined highly effective learning techniques with a machine-learning algorithm that tracks users' forgetting curves and prompts them to review information when they are likely to forget it. Results from this pilot study showed improvements in trainee knowledge retention from the beginning to the end of their course and a reduced risk of failing performance assessments relative to trainees who did not use the app. Furthermore, because the machine-learning algorithm helps increase the efficiency of learning, these positive outcomes were achieved for about one minute 40 seconds of engagement with the app per user, per week. These findings highlight a positive development toward efficient and effective uses of technology to support knowledge retention in military training and, by extension, support complex decision making.

ABOUT THE AUTHOR

Amy M. Smith is the Chief Science Officer at Blank Slate Technologies and holds a PhD in experimental psychology. Her research investigates cognitive approaches to memory optimization with a focus on using technology to strengthen memory against the effects of aging, neurodegenerative disease, and acute stress. Amy has a history of conducting behavioral research with the U.S. Army and in higher education. At Blank Slate, Amy spearheads experimental and product research, some of which has involved one of Blank Slate's clients, the U.S. Air Force.

Leveraging Machine Learning and Cognitive Science to Enhance Knowledge Retention in Air Force Special Warfare Trainees

Amy M. Smith, PhD
Blank Slate Technologies
New Haven, CT
amy@blankslatetech.co

BACKGROUND

Recent years have seen an uptick in the integration of emerging technologies into human performance optimization efforts. Data from wearable tracking devices (e.g., Brown et al., 2021; Butler et al., 2022) and endocrine biomarkers (e.g., Beckner et al., 2021; 2022) are now integrated into military training to provide actionable feedback about health and readiness. Improvements in virtual and augmented reality technology have yielded simulative environments that provide immersive and more abundant training opportunities (e.g., Ahir et al., 2020; Gluck et al., 2020; Pope et al., 2021). Advancements in machine learning and artificial intelligence have boosted the predictive power of algorithms that input human performance metrics and output individual readiness assessments (e.g., Connaboy et al., 2018; Jenkins et al., 2022; Schultebraucks et al., 2021). These efforts showcase the US military's dedication to training personnel with the most innovative technology to face the unique challenges of modern combat.

However, the attention and resources allocated to physical readiness have greatly outweighed those given to cognitive readiness (Herlihy, 2022). While the physical fitness and health of military personnel is essential, physical readiness merely provides the vessel for the mind to work, to solve problems, and to make decisions with accuracy and speed. It is equally essential that the mind is well-trained so that it can properly recall crucial information, actively consider multiple pieces of information at once, and inhibit irrelevant or distracting input signals. In other words, the mind must be ready to engage in all of the cognitive processes that support good decision-making, and this requires targeted training.

Supporting decision-making means supporting knowledge retention. Most people would agree that they would rather have a paramedic treat a life-threatening injury than a car salesman for one critical reason: a paramedic *knows more* about medicine and can thus make a better decision in that moment than the salesman. To facilitate optimal decision-making, knowledge needs to be strong, flexible, and easily retrievable. Without strong memories that are easily accessible, decisions cannot be made quickly or accurately.

In the context of military training, fostering a knowledge base that is strong, flexible, and accessible means more than giving trainees time to study. Research on the science of learning has demonstrated the inefficiency and ineffectiveness of commonly used methods of learning such as reading, highlighting, rote memorization, and repetitive training exercises. Instead, to achieve the most durable memories, robust research supports the use of three learning techniques in tandem: retrieval practice, interleaving, and spacing (Agarwal & Roediger, 2018; Brown et al., 2014; Carpenter et al., 2022). In practice, this involves a trainee practicing recalling the content that they need to know (i.e., retrieval practice), randomly mixing up the order in which they practice recalling their content (i.e., interleaving), and engaging in this practice repeatedly across intervals of days, weeks, and months (i.e., spacing). When learning information, a common method of intermixing these three learning techniques is to review flashcards every few days for a period of time, making sure to shuffle them before each review session.

Of further relevance to military personnel, the ability to recall information is greatly hindered under conditions of acute stress. Stress triggers the release of the hormone cortisol, which binds heavily to brain regions involved in memory retrieval and impairs functioning (Gagnon and Wagner, 2016; Shields et al., 2017). However, research suggests that when people use the learning strategies described above during training, they create memories that can still be retrieved under stress (Smith et al., 2016, 2018). Thus, not only is the trifecta of retrieval practice, interleaving, and spacing the current gold standard for building a strong knowledge base, it is the only known way to prepare one's knowledge base for the anticipated assault of acute stress.

In 2021, the 351st Special Warfare Training Squadron (SWTS) at Kirtland Air Force Base (AFB) in Albuquerque, New Mexico identified the need for improved knowledge and decision-making in their pararescue (PJ) trainees. PJs are combat rescue personnel who are trained to recover civilian and US and allied forces in permissive and non-permissive environments worldwide. They therefore have expertise in air, land, and sea operations. As necessitated by the job, PJs are elite tactical athletes and highly knowledgeable paramedics. In their training at Kirtland, one of the biggest barriers to success in the 25-week PJ course is passing a series of medical performance assessments in which a PJ must properly diagnose and treat a hypothetical illness or injury. For some trainees, these moderately stressful simulations cause them to forget critical information such as the correct sets of symptoms for certain illnesses, proper dosages of medications, and sequences of action for certain treatments. For others, though they can successfully access their knowledge, the cognitive burden of doing so causes them to miss situational threats like distant enemy fire or shouting that is taking place in the simulated rescue scenario. Thus, the trainees face issues with both knowledge retention and ease of recall, rendering them in need of a learning tool that could help automate memory retrieval.

Recognizing the need to improve cognitive acuity in PJ trainees, the 351 SWTS initiated the use of a new approach through a mobile app interface to support trainee knowledge retention. The app incorporates retrieval practice, interleaving, and spacing into a flashcard-style mobile app that aims to help people retain information over periods of months and years. What is most novel about this approach is its additional use of a machine-learning algorithm that tracks individual forgetting curves to provide individualized schedules of spacing for each trainee. After prompting a trainee to engage in self-testing over a period of several days, the algorithm learns that user's rate of forgetting for each item. It then adjusts the trainee's testing schedule to prioritize items that they are more likely to forget. As demonstrated in a peer-reviewed experiment on this app (McHugh et al., 2021), this adaptive feature greatly improves the efficiency of learning such that individuals do not spend unnecessary time testing themselves on information that they already know well. In addition to the memory support provided to users, the app is accompanied by a continuously updated analytics dashboard where leadership can monitor individual and team knowledge levels. These analytics allow instructors to monitor trainee pain-points in real time and facilitate dynamic approaches to training that address knowledge gaps as they arise.

In summary, to ensure that military personnel are maximally prepared to execute their duties, it is essential to intentionally train and support their cognitive readiness in addition to their physical readiness. The 351 SWTS at Kirtland AFB took steps toward improving the cognitive readiness of their PJ trainees in 2022 when they sought out technology that supports knowledge retention through advancements in cognitive science and machine learning. The present paper presents the methods and results from this pilot study, in which three classes of PJ trainees were provided with a mobile app to support retention of their medic training throughout their course.

METHOD

Participants

Participants consisted of 89 male airmen across three classes of PJ trainees at Kirtland AFB. The three classes ran from January 2022 – May 2022, March 2022 – August 2022, and June 2022 – December 2022. Though no additional demographic information was collected, leadership at the 351 SWTS estimated their ages to be between 20 and 35 years old with most ranging from 20 to 25 years old.

Materials

A bank of 52 multiple-choice and ordered-response questions was created by a Master Sergeant who is a PJ and oversees the development of paramedic training at Lackland AFB in San Antonio, TX. All questions were based on the Pararescue Medical Operations Handbook, a textbook that PJs have access to throughout the course. Questions assessed a variety of detailed information about medical diagnosis and treatment protocols. For example, one prompt was, "125mg of ____ should be taken twice/day 24 hours before ascent to prevent acute mountain sickness" with the answer choices Acetazolamide, Spironolactone, Dramamine, and Prednisone.

Procedure

In the week before the start of their 25-week course, trainees attended an in-person onboarding with their lead instructor that equipped them with all the resources they would need throughout the course (e.g., gear, textbooks). During the onboarding, the instructor explained the purpose of the learning app and told trainees how to access it from either their mobile device or a web browser. The instructor also provided trainees with preset login credentials and encouraged them to log in for the first time and explore the app. The instructor told trainees that app usage was not mandatory but encouraged them to use it to support their knowledge retention for their medical training. During the PJ course, six major skills areas are taught in two- to seven-week modules. The medical module is the first in the course and lasts for five weeks. Thus, students were encouraged to use the app from the start of the course until their final evaluation phase at the end.

Seventy-six of the 89 trainees voluntarily used the app throughout the 25-week course. In practice, this amounted to trainees receiving push notifications and/or emails approximately once every one to five days (depending on their individualized schedule of spacing) that reminded them to log in and review the content that was prioritized for them by the machine-learning algorithm. Trainees were limited to reviewing a maximum of 15 questions per day to lighten the burden on their time. Throughout the course, instructors kept a log of each trainee that failed at least one performance assessment in the course and had to retry said assessment. Because the course has an average success rate of 97-98%, instructors were not asked for the purpose of this research to report the frequency of trainees who failed the course.

Throughout each of the three classes of PJs, the lead instructor accessed the analytics dashboard online to track trainee usage of the app and trainee knowledge levels. After the third class graduated, A. S. conducted an in-depth phone interview with the lead instructor. Interview questions were focused around two main topics: (1) the instructor's impressions of the utility of the app and analytics dashboard during the pilot, and (2) the instructor's impressions of the utility of the app and dashboard in future classes of PJs.

Analysis Details

Data Selection
Except for four trainees, all trainees stopped using the app within the first 16 weeks of their course. Thus, all following analyses on app usage were conducted on data that ranged across the first 16 weeks.

Question Scoring
All questions that were answered using the app were scored as either correct or incorrect. Partial credit was not given for partially correct answers (e.g., when a question had two correct answers but only one was selected).

Descriptive Statistics
In the reporting of descriptive statistics below, both mean and median values are reported as measures of central tendency and both standard deviations and ranges are reported as measures of variability. Though means and standard deviations are more commonly reported, all dependent variables in this study were deemed not normally distributed on Shapiro-Wilk tests of normality (all p's $< .001$). Thus, readers may find it more informative to interpret the medians and ranges.

Comparing Two Measurements
Because none of the dependent variables were normally distributed, non-parametric tests were conducted for all comparisons of two measurements. Wilcoxon signed-rank tests were conducted to compare two dependent measures. Mann-Whitney U tests were conducted to compare two independent groups.

Comparing Frequencies
Fisher's Exact Tests were used to compare observation frequencies between two independent groups. These were chosen over traditional Chi Square tests because the sample size was small (N = 76 trainees who used the app), resulting in fewer than 10 observations in some of the cells of the contingency table.

RESULTS

Knowledge Retention

Figure 1 displays the average learning curve for app content across the first 16 weeks of the course. The average percent correct on app questions started at 67.22% on the first day of training and increased to 80.56% by the 16th week of training. The PJs finished the medical module on day 35, which may account for the plateau in learning curves around that time.

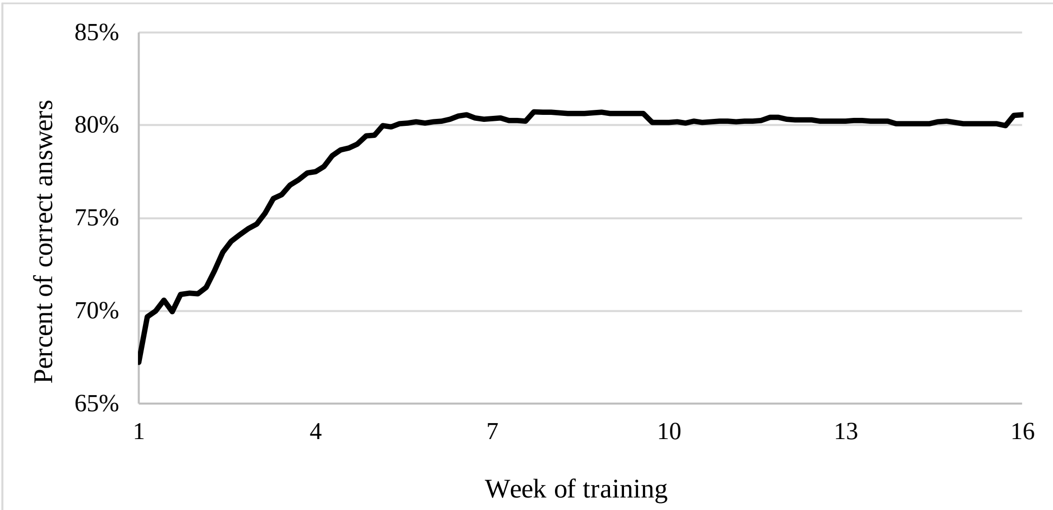

Figure 1. Average Learning Curve for App Content Across All 76 Trainees Throughout the First 16 Weeks of the Course. The Curve Represents the Average Rate at Which Trainees Mastered the 52 Questions in the Question Bank.

The pie charts in Figure 2 display trainee changes in accuracy from the first time they answered all of their 52 questions to the final time they answered all questions. Mean performance across trainees' first answering of all 52 questions was 69.85% (*SD*: 12.88%, *Mdn*: 71.15%, *Range*: 33.33% - 100%). Mean performance on trainee's final answering of all questions was 80.62% (*SD*: 18.26%, *Mdn*: 81.35%, *Range*: 33.33% - 100%).

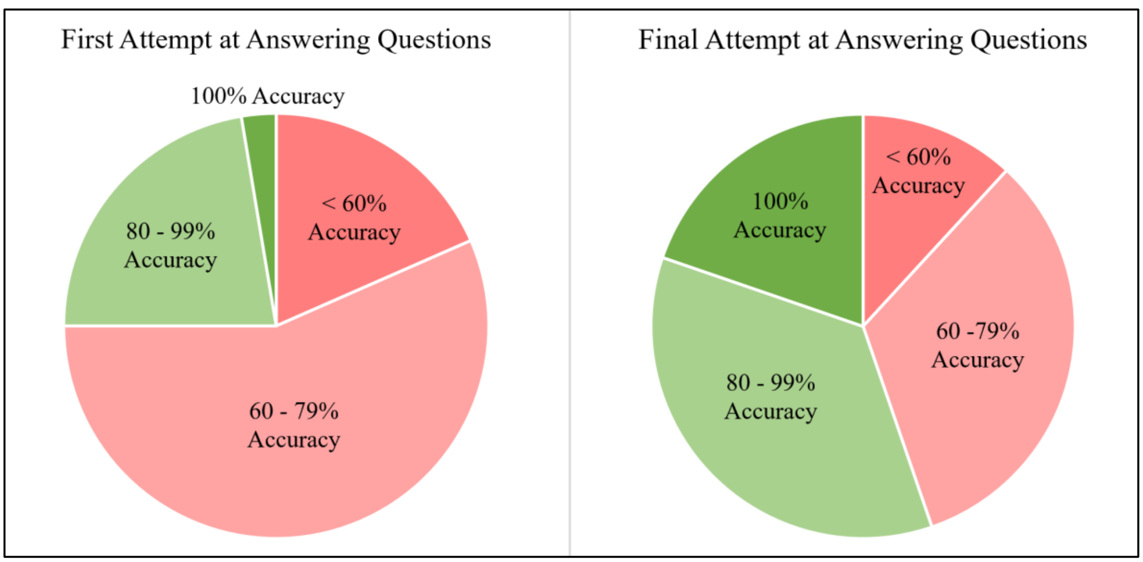

Figure 2. Pie Charts Depicting the Frequencies of Trainees Who Scored in Various Ranges of Accuracy During Their First Attempt at Answering the 52 Questions (left) and Final Attempt at Answering the Questions (right) in the App.

Time Burden

Figure 3 displays how many minutes, on average, each trainee spent per week using the app throughout the first 16 weeks of the course. The PJs finished the medical module on day 35, which may account for the marked decrease in time spent using the app at that point in their training.

In the first three weeks of the course, each trainee averaged between six and nine minutes of total app usage per week. By week 16, each trainee averaged less than one minute of app usage per week. Across all 76 trainees who used the app and all 16 weeks, trainees averaged 100.25 seconds (*SD* = 72.65 seconds) of app usage per week.

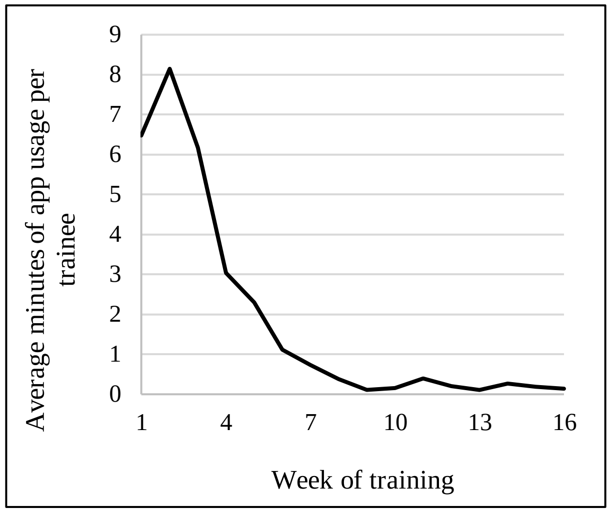

Figure 3. Average Minutes Spent Using the App Per Trainee, Per Week Throughout the First 16 Weeks of the Course.

Performance Assessment Failure and App Use

Of the 89 trainees who were given the app, 76 trainees used it at least one time and 13 trainees never used it. Of the 76 trainees who used the app, 12 failed at least one performance assessment during the course. Of the 13 who did not use the app, 7 failed at least one performance assessment.

A Fisher's exact test was used to determine if there was a significant association between using the app and failing at least one performance assessment. There was a statistically significant association between the two variables (log odds ratio: -1.80, 95% CI [-3.27, -.38], p = .005), suggesting that using the app was associated with a lower likelihood of failing a performance assessment.

Table 1 displays descriptive statistics for app usage metrics for the 76 trainees who did use the app at least one time. Mann-Whitney U tests were conducted to determine whether app usage differed for trainees who did and did not experience performance assessment failure at some point throughout the course. Relative to trainees who did not fail any performance assessments, trainees who failed at least one had a higher total number of sessions (U = 241.50, p = .043), a higher number of total questions reviewed (U = 238.50, p = .039), spent more total minutes using the app (U = 203.00, p = .010), and answered more questions incorrectly (U = 233.00, p = .042). Together, these results show that trainees who failed at least one performance assessment throughout the course were using the app more and were also getting more questions wrong than trainees who did not fail any assessments.

Table 1. Descriptive Statistics for App Usage Metrics for 76 Trainees Who Used the Mobile Application.

	All trainees who used the app (n = 76)	**Trainees who failed one or more performance assessments (n = 12)**	**Trainees who did not fail any performance assessments (n = 64)**
Total number of app sessions			
Mean (*SD*)	11.82 (*10.15*)	16.42 (*9.60*)	10.95 (*10.08*)
Median (*range*)	9 (*1 – 37*)	16 (*4 – 33*)	7 (*1 – 37*)
Total questions reviewed on the app			
Mean (*SD*)	137.32 (*98.59*)	190.08 (*82.75*)	127.42 (*98.72*)
Median (*range*)	126 (*1 – 315*)	216.5 (*67 – 315*)	117.5 (*1 – 306*)
Total minutes spent using the app			
Mean (*SD*)	30.07 (*21.80*)	46.39 (*23.14*)	27.02 (*20.30*)
Median (*range*)	28.44 (*0.15 – 78.25*)	48.39 (*13.97 – 78.25*)	26.46 (*0.15 – 73.65*)
Average number of questions answered incorrectly on the app			
Mean (*SD*)	21.61 (*13.26*)	28.42 (*8.87*)	20.29 (*13.62*)
Median (*range*)	22 (*1 – 57*)	28 (*16 – 44*)	19 (*1 – 57*)

Qualitative Feedback from the Lead Instructor

Utility of the App

The lead instructor commented on four main ways in which the app and analytics dashboard were useful in PJ training.

First, the instructor described the app as a tool that capitalized on the "gray space" in the course: the time that was not already accounted for by training exercises. In other words, the app did not place a huge time burden on trainees and could be completed during off hours or breaks during training when trainees otherwise did not have any commitments.

Second, the instructor appreciated that instructors could use the app to emphasize the content that they believed was most important and/or needed additional review. Whereas the PJ handbook presents all information equally, instructors could choose to more heavily weight trainees' app questions toward content that the instructors wanted to emphasize.

Third, the instructor found it useful to have access to the analytics dashboard for two purposes. First, the instructor used the dashboard to monitor trainee motivation and dedication to the course. For the first time, instructors had a digital record of which trainees were taking advantage of the resources provided to them and which trainees were not. Second, though the instructor did not personally use the dashboard for this reason, the instructor anticipated it would be beneficial for monitoring trainee knowledge levels throughout the course and using that information to flexibly adapt instruction to address trainee pain-points.

Fourth, in the future, the instructor anticipated that this type of tool could replace pen-and-paper testing, saving the schoolhouse the 2-3 days of time each course that it takes to administer and grade paper exams. The instructor reasoned that if trainees could digitally demonstrate knowledge mastery on questions that cover all their course objectives, they would no longer need to engage in pen-and-paper testing on the same topics.

Barriers to Implementation

The lead instructor discussed two current barriers to implementing a retention-support tool like this on a wider scale. First, trainees would need to have access to learning tools on government-owned devices. Without this, trainees could not be required to use them. Second, instructional staff would need to create a much larger question bank that maps onto their specific course objectives, and creating this corpus of questions would be time- and resource-intensive.

DISCUSSION

In the current pilot study, 89 PJ trainees were given access to a mobile app to support their knowledge retention of medical information throughout their 25-week course. To support knowledge retention, the app combined the learning techniques of retrieval practice, interleaving, and spacing with a machine-learning algorithm that tracks individual rates of forgetting. Overall, trainees demonstrated a 10.8 percentage-point increase in their knowledge proficiency for the app content from their first viewing of each question to their last viewing of each question. Further, this improvement was achieved for an average time commitment of about one minute 40 seconds of app usage per trainee per week during their first 16 weeks of the course. Though this pilot was limited by a small sample size and a small question bank of just 52 questions, the results point to promising directions for future work with this population.

When comparing the 76 trainees who used the app to the 13 who chose not to, the analysis showed that the trainees who used the app were less likely to fail a performance assessment during their course. One interpretation of this association is that using the app helped trainees pass their assessments by making their medical knowledge more readily accessible. However, there are alternative explanations for these results. For example, trainees with low motivation and/or poor organizational skills would likely express both behaviors: choosing not to use the app and failing course assessments. Further, because the data regarding specific reasons why trainees failed their respective performance assessments were not available, it could not be determined whether improvements in medical knowledge were driving higher assessment pass rates. To determine a more causal relationship between app usage and passing course assessments, future research with this group must involve systematic measurement of the course outcomes that are directly related to the content that is provided on the app.

In the analyses on app usage patterns for trainees who did and did not fail any performance assessments, trainees who did fail performance assessments used the app more than those who did not. Importantly, these trainees also answered more questions incorrectly when using the app. Taken together, these findings paint a clear picture. Counter to intuition, people who are struggling to learn their content may use an app like this more than people who know their content well. Because machine-learning algorithms of this nature detect forgetting curves and thus only require users to engage when they're likely to forget something, people who consistently answer their questions correctly are reminded less often to review them. This results in less app usage for people who answer their questions with higher accuracy. Like the Yerkes-Dodson law of arousal and performance, there is an inverted U-shaped relationship between amount of app usage and performance. Lower-than-average usage and higher-than-average usage may be associated with low knowledge levels, while a moderate amount of usage yields the optimal results.

The interview with the lead instructor of the PJ schoolhouse further elucidated the potential for learning apps to be useful for instructors and leadership. The instructor appreciated the light-touch nature of the app that was afforded by the machine-learning algorithm, the platform that the app provided for emphasizing the most critical course content, and the potential for digitized knowledge testing to replace paper-and-pencil testing. The instructor also appreciated the analytics dashboard for the enhanced visibility it provided into trainee knowledge and motivation. The instructor acknowledged that, to use a knowledge-retention tool to its fullest, the question bank would need to be much larger and usage would need to be mandated. Future work into the efficacy of long-term knowledge retention applications should address these limitations.

Applications that support long-term knowledge retention take a step toward addressing the cognitive readiness gap in military training. In the current model of military training, trainees are often incentivized to learn information just well enough to pass a test and, in the process of doing so, use poor learning strategies that do not promote memory retention (Herlihy, 2022). Strong and reliable knowledge bases are built over time and spaced, interleaved retrieval practice is an essential tool for building and maintaining memories. Because the consequences of forgetting are steep in the life-or-death scenarios that military personnel face, it is essential that they are provided with tools to support their knowledge. Tools that combine robust learning practices with innovative technologies show promise in this manner: not only as tools to be used throughout military training, but to be used over the course of entire careers.

ACKNOWLEDGEMENTS

I would like to thank the staff and trainees at the 351 SWTS at Kirtland AFB who made this work possible, especially Major Caitlin Harris, CMSgt Michael Rubio, TSgt Ty Hatcher, and Dr. Mark Cantrell.

REFERENCES

Agarwal, & Roediger, H. L. (2018). Lessons for learning: How cognitive psychology informs classroom practice: Laboratory science and classroom observation reveal four simple strategies that can promote learning. (WHAT WE'VE LEARNED ABOUT LEARNING). *Phi Delta Kappan, 100*(4), 8–8.

Ahir, Govani, K., Gajera, R., & Shah, M. (2020). Application on Virtual Reality for Enhanced Education Learning, Military Training and Sports. *Augmented Human Research, 5*(1). https://doi.org/10.1007/s41133-019-0025-2

Beckner, Conkright, W. R., Eagle, S. R., Martin, B. J., Sinnott, A. M., LaGoy, A. D., Proessl, F., Lovalekar, M., Jabloner, L. R., Roma, P. G., Basner, M., Ferrarelli, F., Germain, A., Flanagan, S. D., Connaboy, C., & Nindl, B. C. (2021). Impact of simulated military operational stress on executive function relative to trait resilience, aerobic fitness, and neuroendocrine biomarkers. *Physiology & Behavior, 236*, 113413–. https://doi.org/10.1016/j.physbeh.2021.113413

Beckner, Main, L., Tait, J. L., Martin, B. J., Conkright, W. R., & Nindl, B. C. (2022). Circulating biomarkers associated with performance and resilience during military operational stress. *European Journal of Sport Science, 22*(1), 72–86. https://doi.org/10.1080/17461391.2021.1962983

Brown, P. C., Roediger, H. L., & McDaniel, M. A. (2014). Make it stick: The science of successful learning. Harvard University Press.

Brown, J. A., Whittaker, K. S., Hawkins, S. A., Jackson Santo, T., & Teyhen, D. S. (2021). Using fitness trackers to increase physical activity in a US Army health initiative. *Health Education Journal, 80*(8), 883–896. https://doi.org/10.1177/00178969211026345

Butler, Scott, W. C., Hintz, C. N., Mata, J. D., Bryant, J. F., Tchandja, J. N., Hando, B. R., & Fraser, J. J. (2022). Heat stroke burden and validity of wearable-derived core temperature estimation during elite military training: 1834. *Medicine and Science in Sports and Exercise, 54*(9S), 544–544. https://doi.org/10.1249/01.mss.0000881908.66626.76

Carpenter, Pan, S. C., & Butler, A. C. (2022). The science of effective learning with spacing and retrieval practice. *Nature Reviews Psychology, 1*(9), 496–511. https://doi.org/10.1038/s44159-022-00089-1

Connaboy, Eagle, S. R., Johnson, C., Flanagan, S., Mi, Q., & Nindl, B. C. (2018). Employing Machine Learning to Predict Lower Extremity Injury in U.S. Special Forces. *Medicine and Science in Sports and Exercise*. https://doi.org/10.1249/MSS.0000000000001881

Gagnon, S. A., & Wagner, A. D. (2016). Acute stress and episodic memory retrieval: Neurobiological mechanisms and behavioral consequences. *Annals of the New York Academy of Sciences, 1369*(1), 55-75. https://doi.org/10.1111/nyas.12996

Gluck, Chen, J., & Paul, R. (2020). Artificial Intelligence Assisted Virtual Reality Warfighter Training System. *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 386–389. https://doi.org/10.1109/AIVR50618.2020.00080

Herlihy. (2022). Cognitive Performance Enhancement for Multi-domain Operations. *Parameters (Carlisle, Pa.), 52*(4), 75–. https://doi.org/10.55540/0031-1723.3188

Jenkins, Caballero, W. N., & Hill, R. R. (2022). Predicting success in United States Air Force pilot training using machine learning techniques. *Socio-Economic Planning Sciences, 79*, 101121–. https://doi.org/10.1016/j.seps.2021.101121

McHugh, D., Feinn, R., McIlvenna, J., & Trevithick, M. (2021). A random controlled trial to examine the efficacy of blank slate: A novel spaced retrieval tool with real-time learning analytics. *Education Sciences, 11*(3), 90. https://doi.org/10.3390/educsci11030090

Pope, Ide, J. S., Micovic, D., Diaz, H., Rosenbluth, D., Ritholtz, L., Twedt, J. C., Walker, T. T., Alcedo, K., & Javorsek, D. (2021). Hierarchical Reinforcement Learning for Air-to-Air Combat. https://doi.org/10.48550/arxiv.2105.00990

Schultebraucks, Qian, M., Abu-Amara, D., Dean, K., Laska, E., Siegel, C., Gautam, A., Guffanti, G., Hammamieh, R., Misganaw, B., Mellon, S. H., Wolkowitz, O. M., Blessing, E. M., Etkin, A., Ressler, K. J., Doyle, 3rd, Jett, M., & Marmar, C. R. (2021). Pre-deployment risk factors for PTSD in active-duty personnel deployed to Afghanistan: a machine-learning approach for analyzing multivariate predictors. *Molecular Psychiatry, 26*(9), 5011–5022. https://doi.org/10.1038/s41380-020-0789-2

Shields, G. S., Sazma, M. A., McCullough, A. M., & Yonelinas, A. P. (2017). The effects of acute stress on episodic memory: A meta-analysis and integrative review. *Psychological Bulletin*, *143*(6), 636-675. https://doi.org/10.1037/bul0000100

Smith, A. M., Floerke, V. A., & Thomas, A. K. (2016). Retrieval practice protects memory against acute stress. *Science*, *354*(6315), 1046-1048. https://doi.org/10.1126/science.aah5067

Smith, A.M., Davis, F.C., Thomas, A.K., 2018. Criterial learning is not enough: retrieval practice is necessary for improving post-stress memory accessibility. *Behav. Neurosci. 132* (3), 161–170. https://doi.org/10.1037/bne0000240.