Predicting Learning and Retention in a Complex Task

David Peebles (d.peebles@hud.ac.uk)
Department of Psychology, University of Huddersfield,
Queensgate, Huddersfield, HD1 3DH, UK

Abstract

This paper reports an experiment investigating learning and retention in a complex task over multiple sessions across an extended period of time. The primary aim of the experiment is to evaluate the *Predictive Performance Equation* (PPE: Jastrzembski & Gluck, 2009) a model of learning and forgetting that predicts retention based on past performance. The second aim is to test a taxonomy for knowledge, skills and attitudes and a competence retention analysis technique developed to improve competence retention in military training (Cahillane, Launchbury, MacLean, & Webb, 2013). Participants were trained over 16 weeks on the Multi-Attribute Task Battery (MATB: Comstock Jr & Arnegard, 1992), a computer-based task analogous to piloting an aircraft. The study reveals significant variation in learning profiles for the MATB subtasks and demonstrates the PPE's ability to make accurate predictions of human performance over intervals ranging from 27 to 111 days.

Keywords: MATB; Predictive Performance Equation

Competence retention and training

Many military personnel are required to maintain high levels of task knowledge and skill performance and so are subjected to regimes of regular testing and refresher training to combat the effects of skill fade. The schedule of retraining is typically not determined on an individual basis but is standardised (e.g., calendar-based) and the acceptable threshold criterion is either a general numerical measure such as the number of training hours completed or a qualitative "pass/fail" score. However, because there are substantial differences in people's ability to learn and retain information, it may be the case that two individuals with the same training schedules perform (possibly safety or mission critical) tasks at very different levels of effectiveness.

To complicate matters, there is strong evidence from the psychological literature that knowledge and different types of skills decay at different rates (e.g., Wisher, Sabol, & Ellis, 1999; Stothard & Nicholson, 2001). Together, these two factors suggest that a more efficient and productive approach to training and skill maintenance would be to derive personalised training schedules through detailed analysis of the knowledge, skills and attitudes involved in the task and from each individual's learning and retention profile.

This paper describes an experimental study that aims to investigate and integrate two approaches to the understanding and improvement of competence retention and the personalisation of learning for Defence. The first involves the application of a model of learning and retention called the *Predictive Performance Equation* (PPE)—to create personalised training schedules based on predicted memory retrieval failure (Jastrzembski & Gluck, 2009). The second approach relates to research conducted by the UK Defence Science and Technology Laboratory (Dstl) to develop a set of principles for improving competence retention in military training, together with a *competence retention analysis* (CRA) technique to support competence retention through training (Cahillane et al., 2013). The paper will proceed by first describing the two strands of research, then outlining the details of the experiment, and finally discussing some of the key results, implications and limitations of the study.

The Predictive Performance Equation

The acquisition and retention of knowledge are influenced by three primary factors: the amount of practice (the frequency effect), the amount of time elapsed since the last practice session (the recency effect), and the temporal distribution of practice (the spacing effect). The spacing effect is less intuitive than the others but is a ubiquitous occurrence in learning in which practice sessions which are more widely distributed over time result in better retention compared to identical training sessions scheduled closer together. The beneficial effect of increasing the study interval works only up to a certain point; intervals beyond a certain threshold diminish final retention (Benjamin & Tullis, 2010), but it has been argued that informed use of the spacing effect can have significant positive implications for education and training (Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012).

The PPE characterises the combined effects of recency, frequency and spacing on retention and subsequent task performance (Jastrzembski & Gluck, 2009; Walsh, Gluck, Gunzelmann, Jastrzembski, & Krusmark, 2018). When calibrated to individual task performance data gathered over a series of sessions, the PPE is able to account not only for existing performance data but is also able to make precise, quantitative predictions of an individual's performance at later points in time, sometimes many months into the future (Jastrzembski et al., 2017). It does so by calculating the expected stability of knowledge and skills on the basis of the previous training history and using this measure to predict the expected retention of knowledge and skills across periods of non-use or further practice (Jastrzembski, Portrey, Schreiber, & Gluck, 2013). The PPE is able to provide an accurate estimate of the time when performance has declined to such an extent that refresher training is required.

A key premise of the PPE is that learning new information creates traces in Long-Term Memory (LTM) and that each trace has a degree of *activation*, which determines the probability that it can be subsequently retrieved and the speed with which that retrieval will be accomplished. The mechanism of activation to explain the effect of elapsed time and practice on task performance. According to this account, LTM traces vary in their base level of activation (often referred to as the "strength" of the trace) depending on how frequently or recently they have been used and a trace's strength determines its general availability. The strength of a trace changes gradually; it decays over time but can be progressively increased by repeated practice. The availability of a memory trace is what affects performance.

The PPE equation (Equation 1) predicts the activation of a memory trace, which is subsequently converted to a performance prediction.

$$M_n = (N+a)^c \cdot T^{-d} \tag{1}$$

The PPE assumes that performance increases as a power function (with learning rate c) of the number of practice episodes N, decreases as a power function (with decay rate d) of elapsed time (in seconds) since the episodes occurred T, and that the effects of practice and elapsed time are multiplicative in nature (although, as described above, this effect is on the activation, M of a memory trace, n rather than on performance directly). Finally the a parameter represents an individual's prior experience in the task, adding to the number of practice episodes to increase ease of activation.

In order to incorporate the spacing effect into the equation, a conception of time is introduced in which T is computed as the sum of the time (t_i represents the time of encounter i) since each of the previous study or practice events, each weighted, w_i so that the most recent events are given extra prominence—the older an encounter the smaller the contribution of that encounter to the total time.

$$T = \sum_{i=1}^{n-1} w_i \cdot t_i \tag{2}$$

Weight is an exponential decay function of time according to Equation 3. In this equation, shorter time distances (i.e., more recent encounters with the material) are weighted more heavily, with the x parameter determining the degree of prominence given to shorter time delays. The x parameter is typically set to 0.6 as this provides a good fit to data in many studies.

$$w_i = \frac{t_i^{-x}}{\sum_{j=1}^{n} t_j^{-x}} \tag{3}$$

Equation 4 determines the rate of decay, d for memory traces and is defined to capture the spacing effect in which longer delays between practice episodes result in a reduction in decay rate—and as a consequence produce more stable knowledge (Jastrzembski & Gluck, 2009; Walsh, Gluck, Gunzelmann, Jastrzembski, Krusmark, Myung, et al., 2018).

As the interval between study and test increases, the lower decay rate associated with spaced versus massed repetitions enhances retention. To capture this effect, the function incorporates the history of lags (i.e., time differences) between successive events and is a linear function of the average of one over the sum of the natural logarithm of the lags. In this equation, b and m are parameters that determine the decay intercept and slope of the function and correspond to an individual's overall level of forgetting and their susceptibility to the spacing effect respectively. When lags are long, the value inside the brackets approaches zero, reducing decay to the asymptotic value determined by the b parameter. In contrast, when the lags are short the value inside the brackets approaches one, increasing decay.

$$d_n = b + m \cdot \left(\frac{1}{n-1} \cdot \sum_{j=1}^{n-1} \frac{1}{\log(lag_j + e)} \right) \tag{4}$$

Finally, the level of activation, M_n, computed in Equation 1 is transformed into a continuous response value that represents performance, P_n according to Equation 5. Performance is a logistic (sigmoid) function of activation with range $[0,1]$ where the τ parameter determines the sigmoid's midpoint and the s parameter determines the logistic growth rate (i.e., the steepness of the curve).

$$P_n = \frac{1}{1+\exp\left(\frac{\tau - M_n}{s}\right)} \tag{5}$$

The PPE has been tested in several studies to determine its ability to predict skill fade and when individuals need to return for retraining on critical tasks (Jastrzembski, Gluck, & Rodgers, 2009; Jastrzembski et al., 2013; Gluck et al., 2019; Jastrzembski et al., 2017) and the results so far indicate that the PPE is able to track and predict performance accurately at the individual learner level over timescales ranging from seconds to months.

Competence Retention Analysis

Competence retention analysis (Cahillane et al., 2013) is a novel approach developed by the UK Ministry of Defence (MoD) aimed at formulating a set of generic principles and guidance for the optimisation of competence retention in military training. To achieve this, a new classification of the knowledge, skills and attitudes (KSA) was developed that was consistent with the current psychological literature on mechanisms underlying competence retention and their differential rates of decay.

The primary aim of the CRA is to be a framework grounded in psychological evidence that can provide generic advice and guidance for training designers. Once tasks have been analysed in terms of their cognitive components, designers can consult the CRA to determine the likely retention profiles for the individual components and the task as a whole and then plan refresher training schedules accordingly.

The CRA is based on a three-level categorisation of retention, defined on a criterion value of 50% competence after a given period of time since the last training session. According to this classification, a "high" level of retention is greater than 50% competent after 12 months non-practice, a "moderate" level is 50% competent after 5 months non-practice, and a "low" level of retention is 50% competent after two months non-practice.

In addition, the relationship between psychological components and retention categories can be moderated by the frequency with which they are applied when performing a given task. The CRA defines three frequency levels: "very frequent" (more than once every two months), "moderately frequent" (between once every two months and once every five months), and "infrequent" (once in a period greater than five months). The resulting taxonomy consists of a knowledge domain and four types of skill:

- **Explicit knowledge.** Knowledge required to conduct a task, such as facts, concepts and theories. Retention: High, Frequency: Infrequent.

- **Continuous psychomotor skills.** Tasks requiring the ability to perform well-trained and practiced motor actions that do not have distinct beginnings or endings (e.g., driving, flying an aircraft and target tracking). Retention: High, Frequency: High/Moderate.

- **Discrete psychomotor skills.** Physical tasks with discrete beginnings and endings that rely on both procedural and perceptual motor skills (e.g., disassembling a weapon or other weapon handling tasks). Retention: High, Frequency: High/Moderate.

- **Procedural skills.** Tasks requiring working memory to remember a sequence of steps and their order nature (e.g., using a Battlefield Information Management System (BIMS) to create map overlays (Cahillane & Morin, 2012)). Retention: Low, Frequency: High/Moderate/Infrequent.

- **Decision making skills.** Tasks involving the application of cognitive processes such as, judgement, problem solving and analysis in order for an individual to arrive at a decision (e.g., troubleshooting faulty equipment). Retention: Moderate, Frequency: Infrequent.

Experiment

To reiterate, the experiment has two aims. The first is to determine whether the retention profiles of individuals engaged in a complex task can be captured by the PPE to allow accurate prediction of future performance. The second is to investigate the learning and retention profiles of tasks involving the different psychological domains identified by the CRA.

To achieve both aims, the task selected for the experiment was the Air Force Multi-Attribute Task Battery (AF-MATB; Comstock Jr & Arnegard, 1992; Miller, Schmidt, Estepp, Bowers, & Davis, 2014). MATB is a computer-based interactive multitasking environment consisting of a set of four subtasks designed to be analogous to those performed during aircraft piloting. It has been widely used to study the effects of various factors (e.g., automation, priorities, instructions, task difficulty, etc.) on a range of behavioural measures, including multitasking, attention management, vigilance, decision making, ocular behaviour, prospective memory and subjective mental workload. Crucially for this study, MATB is relevant to Defence and consists of multiple components involving different skill domains where performance can be quantitatively measured. In addition, it has been demonstrated that people learn and improve over time during the task (e.g., Fairclough, Venables, & Tattersall, 2005; Kee et al., 2019).

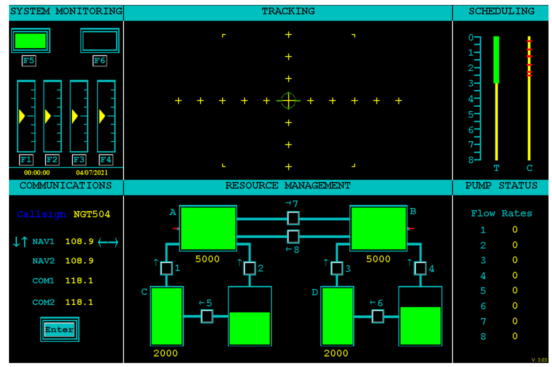


Figure 1: The AF-MATB task interface.

A detailed description of the AF-MATB can be found in Miller et al. (2014) but to summarise, the display consists of four task windows and two information windows (shown in Figure 1). The four tasks and their performance measures are:

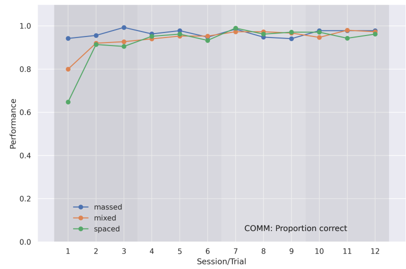
- **System monitoring (SYSMON).** Participants monitor gauges and warning lights and must respond to changes by pressing an appropriate key within a time interval. Performance is measured as the proportion of correct responses.

- **Tracking (TRACK).** Participants must use a joystick to keep a randomly moving cursor inside a target area. Performance is measured as the root mean square deviation (RMSD) distance between the central crosshair and target.

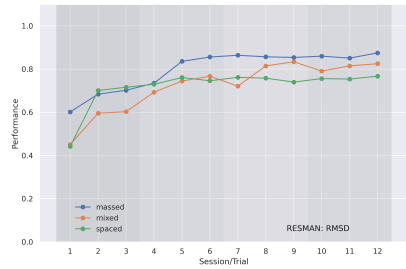
- **Communication (COMM).** Participants must respond to specific auditory messages by adjusting radio and frequency values based on the message. Performance is measured as the proportion of correct adjustments.

- **Resource management (RESMAN).** Participants must maintain the fuel tank levels within target ranges by turning on or off a set of pumps. Performance is measured as the root mean square deviation (RMSD) between actual and target fuel levels.

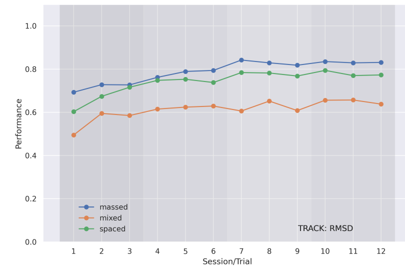
For the purposes of this study, three of the subtasks were associated with CRA skill domains: the TRACK task with the continuous psychomotor (high retention) domain, the RESMAN task with the decision making (moderate retention) domain, and the COMM task with the procedural (low retention) domain. To the extent that these subtasks require the use of a particular CRA cognitive domain, it is expected that their retention profiles will differ. Specifically, the TRACK task should be retained better than the RESMAN task which in turn should be retained better than the COMM task.



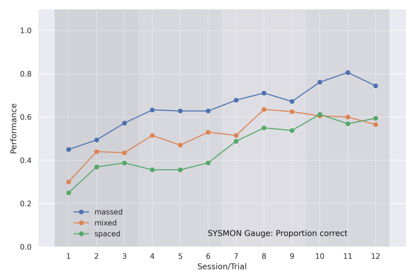
(a) COMM (CM-P)



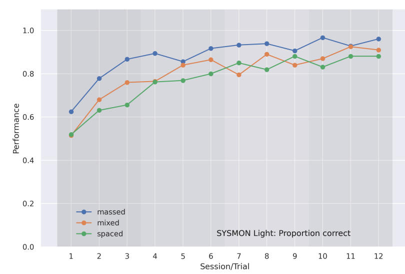
(b) RESMAN (RM-P)



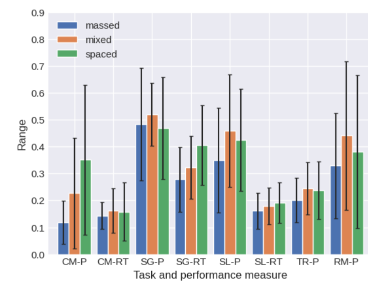
(c) TRACK (TR-P)



(d) SYSMON Gauge (SG-P)



(e) SYSMON Light (SL-P)



(f) Performance ranges

Figure 2: 2a–2e: Learning profiles for the three training schedule conditions over the four training sessions. 2f: Range of performance (P) and task completion times (RT) values for each subtask. Error bars indicate standard deviation.

Participants and materials

Participants were 27 staff, faculty and students from the University of Huddersfield. All participants were 18 years old or over, had normal or corrected to normal eyesight, and were paid £10 per session. The experiment was conducted on PCs running Microsoft Windows 10 with 24-inch displays at 1080p resolution. Participants interacted through the computer keyboard and a Logitech G Extreme 3D PRO joystick.

Design and procedure

The experiment lasted 16 weeks and consisted of four training sessions followed by two assessment sessions[1]. There were three training schedule conditions. Participants in the "massed" condition (9 in total) attended once a day for 4 consecutive days in Week 4. Participants in the "spaced" condition (8 in total) attended once a week (on the same day) for 4 consecutive weeks, while participants in the "mixed" condition (10 in total) attended twice a week for 2 alternate weeks.

For all conditions, the fifth and sixth testing sessions were approximately 42 days (Week 10) and 84 days (Week 16) after the last training session. The aim of creating different scheduling conditions was to provide variation in training spacing as this is a key determinant of the PPE model's predictions. Because the spacing effect is a very well-established result however (e.g., Latimier, Peyre, & Ramus, 2021), this experiment was not designed to include statistical analysis of any effects of spacing on human performance.

MATB event schedules were created to define three 10-minute trials which were different in terms of their event scheduling but equal in difficulty (i.e., number of events and the degree of multitasking required to process them). The trials were designed to be challenging to enable continued performance improvements over the sessions and minimise the likelihood of performance reaching ceiling. A previous MATB study of learning found that sustained learning was only observed in trials where the task demand was high (Fairclough et al., 2005).

Before the first session, participants were introduced to the MATB task via a video which explained the four subtasks and how to interact with the software. Training and assessment sessions consisted of three 10-minute trials (in random order) separated by rest intervals of up to 5 minutes. After each trial, the experimenter would restart the software and ensure that participants had a break and were ready to continue.

Results

Space limitations preclude a full account of the analyses here but further details and data are provided on the study's OSF web page. The sections below will first describe the results of applying the PPE to predicting individual performance in the two test sessions and then report the learning profiles for the MATB subtasks over the four learning sessions.

PPE predictions of human performance While it is possible to apply the PPE to predict performance on the individ-

[1]The study was preregistered with the Open Science Framework (osf.io/uc4fy) and was approved by the Research Ethics Committees of the Ministry of Defence and the University of Huddersfield School of Human and Health Sciences

Table 1: Comparison between human performance and model predictions, sessions 5 and 6

		Session 5						Session 6					
P	Schedule	Human	Model	R^2	RMSE	Days	4-5 Diff	Human	Model	R^2	RMSE	Days	5-6 Diff
1	Mixed	0.654	0.644	0.931	0.024	42	-0.031	0.627	0.618	0.881	0.031	42	0.016
2	Mixed	0.620	0.581	0.887	0.032	42	0.003	0.623	0.580	0.857	0.036	42	-0.022
3	Mixed	0.645	0.636	0.155	0.034	43	-0.059	0.583	0.601	0.010	0.052	48	-0.071
4	Mixed	0.568	0.589	0.518	0.026	42	0.004	0.583	0.589	0.428	0.028	42	-0.011
5	Mixed	0.613	0.672	0.959	0.019	42	0.117	0.624	0.662	0.938	0.023	43	0.100
6	Mixed	0.856	0.846	0.949	0.008	63	-0.036	0.855	0.844	0.895	0.012	28	0.000
7	Spaced	0.553	0.610	0.920	0.025	42	0.084	0.545	0.585	0.842	0.030	42	0.015
8	Spaced	0.495	0.620	0.923	0.017	42	0.080	0.543	0.553	0.695	0.031	42	0.018
9	Mixed	0.673	0.735	0.932	0.019	42	0.035	0.694	0.702	0.822	0.028	41	0.007
10	Mixed	0.702	0.695	0.949	0.021	41	-0.037	0.693	0.688	0.934	0.023	43	0.024
11	Mixed	0.754	0.751	0.863	0.016	41	-0.026	0.760	0.734	0.720	0.021	43	-0.051
12	Spaced	0.729	0.731	0.918	0.025	56	-0.002	0.746	0.712	0.847	0.034	27	-0.014
13	Spaced	0.772	0.809	0.844	0.016	42	0.028	0.781	0.788	0.724	0.020	42	0.028
14	Mixed	0.719	0.728	0.911	0.020	41	0.071	0.778	0.721	0.887	0.022	43	-0.031
15	Spaced	0.761	0.769	0.970	0.014	42	0.026	0.757	0.761	0.943	0.019	42	0.030
16	Spaced	0.797	0.777	0.954	0.014	42	0.007	0.749	0.769	0.913	0.019	42	0.033
17	Spaced	0.820	0.746	0.512	0.026	42	-0.111	0.786	0.752	0.431	0.031	42	-0.022
18	Spaced	0.552	0.576	0.855	0.027	42	0.003	0.551	0.568	0.823	0.030	42	0.036
19	Massed	0.756	0.732	0.877	0.021	50	-0.043	0.708	0.728	0.857	0.022	41	0.041
20	Massed	0.786	0.784	0.623	0.020	42	0.026	0.790	0.787	0.648	0.020	42	0.026
21	Massed	0.679	0.775	0.896	0.025	42	0.087	0.734	0.711	0.740	0.036	42	-0.036
22	Massed	0.783	0.803	0.946	0.017	49	0.008	0.778	0.775	0.916	0.020	35	0.005
23	Massed	0.605	0.666	0.899	0.027	47	0.044	0.615	0.616	0.846	0.032	37	0.012
24	Massed	0.784	0.803	0.738	0.011	42	0.007	0.809	0.794	0.723	0.011	42	0.000
25	Massed	0.748	0.769	0.834	0.026	42	0.041	0.729	0.750	0.817	0.026	42	0.063
26	Massed	0.785	0.811	0.693	0.015	42	0.027	0.791	0.806	0.664	0.016	35	0.051
27	Massed	0.771	0.787	0.935	0.015	35	0.004	0.756	0.743	0.838	0.023	111	0.002
	Mean	0.703	0.720	0.829	0.021	43.7	0.013	0.703	0.701	0.764	0.026	43.1	0.009
	StDev	0.095	0.080	0.186	0.006	5.4	0.050	0.091	0.084	0.202	0.009	14.3	0.036

ual MATB subtasks, for this study the subtask measures were transformed onto a common scale and then averaged for each participant to create a single, global MATB score.

The PPE's predictions were tested on sessions 5 and 6, approximately 43 and 86 days respectively after a participant's fourth training session. For each test, the model was fitted to the individual's performance data from the previous sessions by adjusting five free parameters: b and m, representing the intercept and slope of the decay function (Equation 4) respectively, τ and s which determine the intercept and slope of the activation transformation function (Equation 5) respectively, and a representing an individual's prior experience (Equation 1). The fitted model was then used to predict performance at the date and time of the first trial of the test session.

Table 1 displays the results of the modelling, showing participants' performance, model predictions, and the difference in participants' performance from the last trial of the previous session and the first trial of the current session. Participants' performance varies widely in both sessions (e.g., compare participants 6 and 8) but there was typically little change in performance between sessions, despite a mean interval of approximately 43 days, indicating that, in general, retention remained stable. With a few notable exceptions (e.g., participants 3 and 17 who showed little decay in performance or, somewhat counterintuitively, performance improvements, after time delays), the PPE was able to provide a close fit to the data and make accurate predictions beyond the training set.

Subtask learning profiles Figure 2 shows the learning profiles for the MATB subtasks over the 12 trials of the four learning sessions. All performance measures are scaled to the range [0,1] to allow comparison. Figure 2f depicts the range of performance scores and task completion times produced by each schedule condition for the different subtasks. For example, the low performance range for the TRACK task reflects the relatively shallow learning curves, in contrast to the much greater changes found in the SYSMON gauge task.

Although the relatively small number of participants limits comparison of the schedule conditions, interesting features can seen in the individual subtask data for all three. First, there was a general level of consistency in performance between the three training schedules, not only in the ranges of values produced but also in the performance profiles across the training phase. Performance differences were also evident in the four subtasks. For example, participants in all three conditions quickly achieved and maintained very high levels of accuracy in the COMM task, whereas in the RESMAN task, performance increased more gradually by approximately 30% to 40% during the course of training.

These differences are likely due to the nature of the interactions required. For example, the time constraints of the COMM task demand immediate attention and a rapid sequence of actions to encode, retain, and then enter information into the system, a task that participants cannot complete much faster than 4.5 seconds. Performance improvements in the other subtasks are likely to be due to, amongst other things, the refinement of local and global strategies, for example revising priorities when balancing different resources in the RESMAN task and more efficiently allocating attention when managing competing demands from subtasks.

Discussion

This experiment has generated a rich dataset of individual learning and forgetting in a complex task involving multiple sub-tasks which is yet to be fully analysed. The main analysis reported here however provides additional support for the PPE by demonstrating its ability to predict performance accurately over retention intervals ranging from 27 to 111 days. While the limited number of participants precludes rigorous statistical analysis of the training schedule conditions, the different subtask learning profiles do provide valuable initial pointers for further investigation. While the pattern of differences in learning are not consistent with the classification provided by the CRA, additional analysis of the data from sessions 5 and 6, combined with a detailed task analysis, may provide further insight into differences in retention over longer intervals.

Acknowledgments

This work was funded by the Defence Science and Technology Laboratory (Dstl) on behalf of Ministry of Defence (MOD). The author would like to acknowledge the support and contributions of Dstl, the Human Social Science Research Capability (HSSRC), MOD Stakeholders, participants in the experiment and Tiffany Jastrzembski and Michael Krusmark for their help and advice in developing the model.

References

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology*, *61*(3), 228–247.

Cahillane, M. A., Launchbury, C., MacLean, P., & Webb, S. (2013). *Competence retention* (Tech. Rep. No. DHC-STC_12_T_T2_001_1.1/005 V5.0). Defence Science and Technology Laboratory.

Cahillane, M. A., & Morin, C. (2012). Skills retention in a complex battlefield management system: A pilot study. *Journal of Battlefield Technology*, *15*(1), 65.

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, *24*(3), 369–378.

Comstock Jr, J. R., & Arnegard, R. J. (1992). *The multi-attribute task battery for human operator workload and strategic behavior research* (Tech. Rep. No. NASA-TM-104174). National Aeronautics and Space Administration.

Fairclough, S. H., Venables, L., & Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology*, *56*(2), 171–184.

Gluck, K. A., Collins, M., Krusmark, M., Sense, F., Maaß, S., & van Rijn, H. (2019). Predicting performance in cardiopulmonary resuscitation. In *Proceedings of the 17th international conference on cognitive modelling. ICCM 2019.* Montreal, Canada.

Jastrzembski, T. S., & Gluck, K. A. (2009). A formal comparison of model variants for performance prediction. In A. Howes, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 9th international conference on cognitive modeling.* Manchester, UK.

Jastrzembski, T. S., Gluck, K. A., & Rodgers, S. M. (2009). The predictive performance optimizer: An adaptive analysis cognitive tool for performance prediction. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 53, pp. 1642–1646).

Jastrzembski, T. S., Portrey, A. M., Schreiber, B. T., & Gluck, K. A. (2013). Improving military readiness: Evaluation and prediction of performance to optimize training effectiveness. In W. Arthur, Jr., E. A. Day, W. Bennett, Jr., & A. M. Portrey (Eds.), *Individual and team skill decay* (pp. 177–199). Routledge.

Jastrzembski, T. S., Walsh, M. M., Krusmark, M., Kardong-Edgren, S., Oermann, M., Dufour, K., ... Stefanidis, D. (2017). Personalizing training to acquire and sustain competence through use of a cognitive model. In D. Schmorrow & C. Fidopiastis (Eds.), *Augmented cognition. Enhancing cognition and behavior in complex human environments. AC 2017* (Vol. 10285). Springer.

Kee, T., Weiyan, C., Blasiak, A., Wang, P., Chong, J. K., Chen, J., ... Asplund, C. L. (2019). Harnessing CURATE.AI as a digital therapeutics platform by identifying N-of-1 learning trajectory profiles. *Advanced Therapeutics*, *2*(9), 1900023.

Latimier, A., Peyre, H., & Ramus, F. (2021). A meta-analytic review of the benefit of spacing out retrieval practice episodes on retention. *Educational Psychology Review*, *33*, 959–987.

Miller, W. D., Schmidt, K. D., Estepp, J. R., Bowers, M., & Davis, I. (2014). *An updated version of the US Air Force multi-attribute task battery (AF-MATB)* (Tech. Rep. No. AFRL-RH-WP-SR-2014-0001). Air Force Research Laboratory.

Stothard, C., & Nicholson, R. (2001). *Skill acquisition and retention in training: DSTO support to the army ammunition study* (Tech. Rep. No. DSTO-CR-0218). Electronics and Surveillance Research Laboratory, Defence Science & Technology Organisation.

Walsh, M. M., Gluck, K. A., Gunzelmann, G., Jastrzembski, T., Krusmark, M., Myung, J. I., ... Zhou, R. (2018). Mechanisms underlying the spacing effect in learning: A comparison of three computational models. *Journal of Experimental Psychology: General*, *147*(9), 1325–1348.

Walsh, M. M., Gluck, K. A., Gunzelmann, G., Jastrzembski, T. S., & Krusmark, M. (2018). Evaluating the theoretic adequacy and applied potential of computational models of the spacing effect. *Cognitive Science*, *42*, 644–691.

Wisher, R. A., Sabol, M. A., & Ellis, J. A. (1999). *Staying sharp: Retention of military knowledge and skills* (Tech. Rep. No. 39). US Army Research Institute for the Behavioral and Social Sciences.