Jastrzembski, T. S., Gluck, K. A., & Gunzelmann, G. (2006). Knowledge tracing and prediction of future trainee performance. In *Proceedings of the 2006 Interservice/Industry Training, Simulation, and Education Conference* (pp. 1498-1508). Orlando, FL: National Training Systems Association.

Knowledge Tracing and Prediction of Future Trainee Performance

Tiffany S. Jastrzembski
Florida State University
Tallahassee, Florida
jastrzembski@psy.fsu.edu

Kevin A. Gluck, Glenn Gunzelmann
Air Force Research Laboratory
Mesa, AZ
kevin.gluck@mesa.afmc.af.mil,
glenn.gunzelmann@mesa.afmc.af.mil

ABSTRACT

Intelligent tutoring systems seek to optimize instruction and training by adapting and individualizing the learning experience on the basis of a student model (Shute, 1995). This model represents the system's estimate of the student's current knowledge or skill level, established from a performance history. Knowledge tracing (Aleven & Koedinger, 2002; Anderson, Conrad, & Corbett, 1989) is a dynamic, Bayesian approach to updating the estimates of probability of skill mastery in the student model. A fundamental shortcoming of this approach is that it does not include a representation of memory decay during periods of non-practice. As a result, traditional student modeling approaches are unable to make predictions regarding knowledge and skill changes under various future training schedules or to prescribe how much training will be required to achieve specific levels of readiness at a specific future time. In this paper, we propose a new knowledge tracing equation, computationally inspired by the learning and forgetting equations in the ACT-R cognitive architecture (Anderson et al., 2004), which uses performance history to baseline student model parameters and then extrapolates knowledge state transformation to predict future performance. We explore practical issues concerning predictive models of future trainee performance and the prescription of frequency and timing of optimal learning with training systems. For instance, we investigate how much data from the training history are necessary to achieve reasonable predictive validity, and we describe the impact of data granularity through a quantitative assessment of how adequately the model can fit and predict human performance curves across aggregate-level, team-level, and individual-level resolutions. The paper ends with a discussion of the implications of this research for the future of training and education.

ABOUT THE AUTHORS

Tiffany S. Jastrzembski is a doctoral student of Cognitive Psychology at Florida State University. Her research interests include computational and mathematical modeling, expertise, cognitive aging, human factors and human engineering. She completed the research described in this paper as a summer intern at the Air Force Research Laboratory's Warfighter Readiness Research Division in 2005, an appointment administered through the Oak Ridge Institute for Science and Education. She earned a B.S. (1998) in Psychology from Carnegie Mellon University and an M.S. (2003) in Cognitive Psychology from Florida State University, with a PhD in Cognitive Psychology from Florida State University expected in 2006.

Kevin A. Gluck is a Senior Research Psychologist at the Air Force Research Laboratory's Warfighter Readiness Research Division, where he leads the Performance and Learning Models (PALM) research team. Current PALM team research areas include cognitive orientation, visuospatial working memory, language-enabled agents, learning and forgetting, cognitive moderators, and Predator pilot modeling. Kevin earned his bachelor's degree at Trinity University in 1993 and his PhD in Cognitive Psychology from Carnegie Mellon University in 1999.

Glenn Gunzelmann is a Research Psychologist at the Air Force Research Laboratory's Warfighter Readiness Research Division. He earned his bachelor's degree in psychology at Albright College (1997), a master's degree in Cognitive Psychology from the University of Florida (1999), and a Ph.D. in Cognitive Psychology from Carnegie Mellon University in 2003. His main research interests focus on developing computational cognitive models of the mechanisms of human spatial competence and the impact of cognitive moderators on human performance.

Knowledge Tracing and Prediction of Future Trainee Performance

Tiffany S. Jastrzembski
Florida State University
Tallahassee, Florida
jastrzembski@psy.fsu.edu

Kevin A. Gluck, Glenn Gunzelmann
Air Force Research Laboratory
Mesa, AZ
kevin.gluck@mesa.afmc.af.mil,
glenn.gunzelmann@mesa.afmc.af.mil

INTRODUCTION

Intelligent tutoring systems are intended to optimize learning by adapting training experiences on the basis of proficiency. These systems continuously estimate trainees' current knowledge and skill levels based on performance history and build what has been termed a *representation of the student* (Hartley & Sleeman, 1973) or *student model* (Greer & McCalla, 1993; Shute & Psotka, 1996; VanLehn, 1988). They dynamically update estimates of the knowledge state in the student model as the learner accumulates more experience and expertise, and then adapt training to improve the efficiency and effectiveness of learning opportunities.

Among the most demonstrably successful intelligent tutoring systems ever created are the Cognitive Tutors® that originated at Carnegie Mellon as testbeds for the ACT* theory of skill acquisition (Anderson, 1983). Their implementation was inspired by ACT-style cognitive models of algebra and geometry problem solving, with skills decomposed into production rules. The tutors proved so effective that a successful spinoff company, Carnegie Learning, eventually formed to mature and distribute the technology to school districts around the country. The tutors are now being used by more than 800 schools.

The student modeling capability in the Cognitive Tutors® is a Bayesian estimate of the probability of having mastered each of the knowledge units (production rules) that are targets of current instruction. Their Bayesian equation is used in a process called *knowledge tracing* (Corbett & Anderson, 1995) to keep this mastery estimate current and provide a basis on which to determine the course of instruction. This approach has been quite successful in classroom applications. (Aleven & Koedinger, 2002; Anderson, Conrad, & Corbett, 1989).

Notwithstanding the documented utility of the knowledge tracing approach, it does have a critical limitation, as does every other known student modeling approach. The limitation is that intelligent tutors have no underlying mechanism for memory decay represented in the model. Thus, even over significant periods of non-practice, when some forgetting would inevitably occur, the student model assumes that the learner's knowledge state remains stable across periods of non-use, leaving all prior learning completely intact. This limits the utility of traditional student modeling approaches entirely to estimates of *current* readiness/proficiency/mastery. They have no capacity to predict what *future* readiness will be at specific points in time.

Furthermore, traditional student modeling approaches are unable to make predictions regarding knowledge and skill changes under various future training schedules or to prescribe how much training will be required to achieve specific levels of readiness at a specific future time. They function only on the learner's last computed knowledge state, and provide training for only the current benchmark task needed to be learned.

The goal of the current work is to further translate basic cognitive science research into an effective "cognitive tool" (Koedinger & Anderson, 1993) for future warfighter training applications. We will do this through the creation of a mathematical model that integrates mechanisms that handle the spacing effect (distributed learning) into a computational cognitive process model of memory. Benefits associated with computationally representing the spacing effect include validating existing or proposed theoretical assumptions of learning and decay of memory traces over time, providing warfighters and instructors with a tool to predict performance given a known regimen of training, and helping warfighters and instructors prescribe practice schedules to optimize performance based upon mathematical regularities in training histories.

We propose a new knowledge tracing equation, inspired largely by the learning and forgetting equations in the ACT-R cognitive architecture (Anderson et al., 2004). This equation allows us to calibrate student model parameters from performance history and extrapolate knowledge state transformation to predict future performance. We first begin with an explanation of the spacing effect dilemma, then turn to the evolution of computational models to formally trace the intricacies of knowledge and skill acquisition in human memory. Finally, we address the potential contributions of a predictive and prescriptive cognitive model for improving military readiness.

SPACING EFFECT

One of the most consistent findings from past research in human memory is that performance is generally enhanced when learning repetitions are spaced farther apart temporally. This phenomenon, often termed the spacing effect, is extremely robust and has been observed not only in artificial laboratory settings, but in real-life training situations as well (e.g. Bahrick & Phelps, 1987). Due to its ubiquity, it may be inferred that basic principles of learning and retrieval are involved.

On the learning side of the coin, practice that occurs more slowly becomes more durable (e.g. Pavlik & Anderson, 2005); and on the forgetting side of the coin, the rate of forgetting of an item decreases as time passes according to Jost's Law. This Law states that "if two associations are now of equal strengths but of different ages, the older one will lose strength more slowly with the further passage of time" (Woodworth, 1938).

This phenomenon is not captured by most existing models of human memory, which generally assume that memory traces additively strengthen with each learning opportunity and continually decay with the passage of time. Thus, computational models fall apart under distributed training conditions and it becomes evident that modifications to current implementations of computational models of memory need to be made to account for differences in learning and decay as a function of repetition timing.

COGNITIVE MODELS

Computational cognitive process models have been in existence a mere fraction of the hundred and twenty years of accrued research in human learning and forgetting of knowledge and skill (Ebbinghaus, 1885). Despite their infancy, such models have capitalized on theoretical and empirical understandings to inform the mathematical implementation of cognitive mechanisms and processes responsible for performance. Significant strides have been made in accounting for increasingly complex memory phenomena through the years (e.g. Anderson, 1992; Anderson & Lebiere, 1998; Anderson, Fincham, & Douglass, 1999; Pavlik & Anderson, 2005). However, much work remains to be done to completely capture the nuances of the dynamic human memory system. As it currently stands, even the best models in existence capture learning and forgetting curves only in a post-hoc manner, adequately simulate curves only when the grain of resolution is large enough to diminish inherent noise and variation and typically account for performance curves averaged over many participants rather than tracing the knowledge state of an individual learner.

ACT-R General Performance Equation

Anderson and Schunn (2000) proposed the General Performance Equation, which provides the basis for our predictive and prescriptive mathematical model. It is derived from ACT-R equations and comprises the power law of practice, the power law of forgetting, and the multiplicative effect of practice and retention (the relation between the amount of practice and the duration of time for which knowledge must be maintained). A form of neural adaptation called *long-term potentiation* also shows the power laws of learning and forgetting (Barnes, 1979), which nicely aligns the cognitive mechanisms of the model with neurophysiological research.

The General Performance Equation is formally expressed as (see Equation 1):

$$A \cdot N^{c} \cdot T^{-d} \qquad (1)$$

where A is a free parameter scalar, N is the amount of practice, c is the rate of learning, T is the time since learning, and d represents memory decay. The collective effect of this algorithm is that performance continues to improve with increased learning opportunities, and continues to degrade as time between learning and retention increases. Preservation of knowledge then depends upon leveraging the amount of practice against the retention time.

To emphasize the reasons for utilizing these core components in our proposed modified equation, we first demonstrate the model's strengths. This ACT-R-based General Performance Equation can replicate the findings from a variety of learning and forgetting studies in the published literature. These include studies concerning knowledge retention, knowledge acquisition, skill retention, and skill acquisition. We provide a sample of these model fits in Figure 1 for knowledge acquisition, and Figure 2 for skill retention.

Anderson and Fincham (1994) required participants to first memorize a number of logic-based facts. These facts related time between series of events, and participants were asked to predict when one event would occur, given the knowledge of when a second event occurred. Participants were tested over the course of four days.

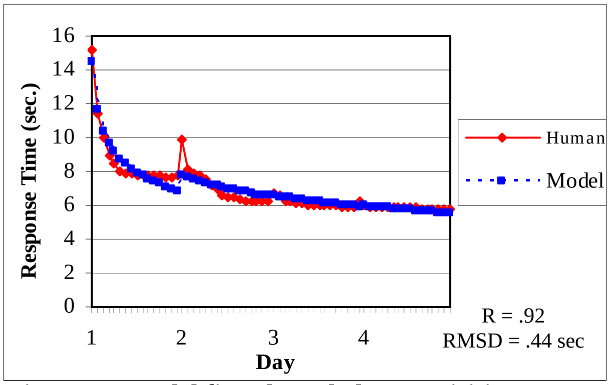

Figure 1: Model fit to knowledge acquisition (Anderson & Fincham, 1994)

Bean (1912) taught novice participants typewriting skills and was interested in examining how well those new skills were retained as a function of time. Participants were initially tested on days one, four, and seven and were then tested weekly for four additional weeks, and tested a final time 35 days after initial learning.

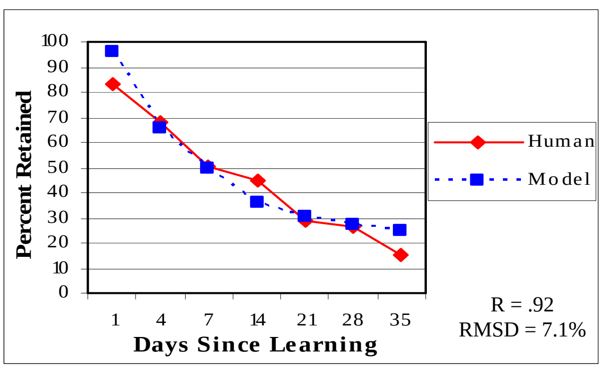

Figure 2: Model fit to skill retention (Bean, 1912)

These figures demonstrate the usefulness of the General Performance Equation for many types of data sets and provide correlation coefficients of 0.89 to 0.97 for fits to empirical human performance. We now turn to a dimension of learning and forgetting that this equation does not handle well, namely, distributed learning or spaced practice.

Mathematical Weaknesses of the General Performance Equation for Handling the Spacing Effect Human performance studies have revealed that learning and forgetting do not linearly improve or degrade over extended periods of time, but rather they approach asymptote. For example, an item presented at longer intervals of time will be retained better than an item crammed more tightly together in temporal space. The practice function in its current form would assume a discrete increment in learning or activation to be added at each presentation time of the item and would necessitate a greater decay rate to be incorporated for an item presented across greater intervals of time. Thus, the General Performance Equation would model superior performance for massed study compared to distributed study, resulting in a converse effect to that of actual human performance. As demonstrated in Figure 3, the model clearly loses its ability to fit human performance data when distributed training regimens are a part of the procedure, and correlations plummet to 0.49. Further, these estimations of fit can only be made in a post-hoc manner.

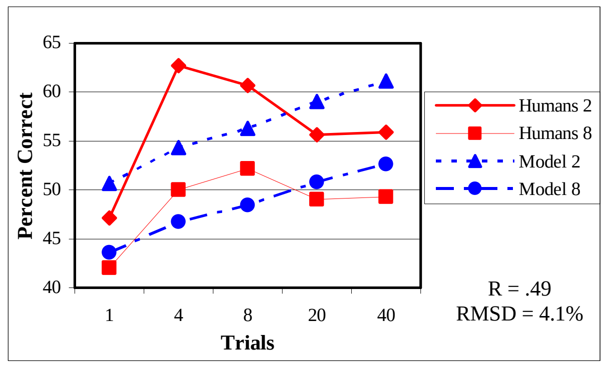

Figure 3: General Performance Equation Model fits to data spaced at practice intervals of every 2 and every 8 trials (Glenberg, 1976)

PROPOSED PREDICTIVE AND PRESCRIPTIVE MODEL

Algorithm Parameters

Building upon the strengths of the previous equations, we sought to formalize an algorithm to capture recency, frequency, and spacing effects, while also providing flexibility and capability for predicting performance at later points in time. This equation is formalized by the following, and incorporates the same definitions for parameters *N* and *c* as originally defined by Equation 1 (see Equation 2):

$$S \cdot N^{c} \cdot T^{-a} \qquad (2)$$

where *S* equals the original scalar (*A)* in the General Performance Equation) multiplied by training history (known improvement rate between initial time of learning and last known retention session), and *a* equals an activation-based decay parameter that enfolds an exponential function into the decay rate (see Equation 3), such that:

$$a = d \cdot e^{(m-1)} + d(intercept) \quad (3)$$

To further elaborate the activation-based decay parameter *a*, *m* equals the activation level at the latest known data point, defined by $ln(T^{-d})$, so that this parameter is calculated from the known training history and is based upon the original decay rate and activation level at the last known point.

Ability to Account for Spacing Effect

In order to demonstrate the efficacy of our Predictive and Prescriptive Model in comparison to the General Performance Equation, we plotted our model fit to the same data set. Figure 4 reveals correlations of 0.96 between our model and the data, showing a marked improvement over the General Performance Equation (r = 0.49).

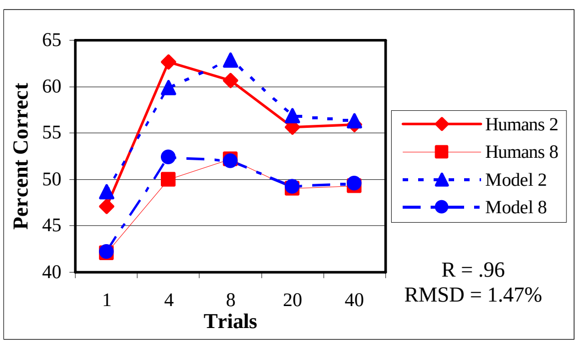

Figure 4: Predictive Performance Equation Model fits to data spaced at practice intervals of every 2 and every 8 trials (Glenberg, 1976)

As we have demonstrated the model's ability to capture recency, frequency, and spacing effects of human memory, we next turn to address its predictive capability utilizing data collected by the Cognitive Engineering Research Institute (CERI) investigating team training and performance.

PREDICTIVE MODEL FITS TO TESTBED DATA

CERI studied human performance in an Uninhabited Air Vehicle (UAV) synthetic reconnaissance task environment, and the data proved to be ideal for examining the accuracy of our model's predictions. In addition, the study design allowed us to investigate model fits at various levels of data resolution, meaning we were able to examine model predictions at the aggregate, team level, and individual team member level of performance.

To provide some background regarding CERI's study design, individual teams were composed of three members randomly assigned to positions (a mission coordinator/route planner, an air vehicle operator (AVO) responsible for piloting the aircraft, and a payload operator (PLO) to operate the camera and take pictures of required targets), and each team member was assigned certain unique duties that provided access to different pieces of information (e.g. the mission coordinator knew the location of targets and airfield restrictions, the altitude/speed technician knew the optimal parameters for reconnaissance photos, and the photographer knew when target reconnaissance was complete so that the aircraft could move onto its next target). Teams were required to work cooperatively so that mission-critical information could be passed along to the appropriate team member to ensure success.

Participants completed five, 40-minute missions on the first day of training and returned 10-14 weeks later to complete three final missions. Outcome measures were based upon weighted penalty scores across team members, amassed across all occurrences of team members acting outside duty restrictions or failing to relay mission-critical information to the appropriate team member. This training scenario will be utilized as the model's baseline of training history for both predictive and prescriptive scenarios described below.

Predictive Restrictions of Computational Models

As predictive capability of any model is affected by the level of noise in the data set, performance trends, and ultimately mathematical regularities, may be difficult to extract if the amount of noise is too high. The model may therefore function according to an inadequate, baseline training history, and may make increasingly poor predictions for future performance as the level of noise rises.

This issue was important to understand as we sought to investigate model fits across finer and finer grains of data analysis. Decomposing the data from the aggregate level downward inherently confounds the identification of true, stable memory gains and losses in performance history (Estes, 2002), since outlier trials, participants, or extraneous error are less likely to be reduced through averaging into the overall trends.

Nonetheless, these examinations will help serve some very practical purposes. They will reveal how much data, at a minimum, is necessary to make valid predictions for individuals or teams performing a given task. These analyses may provide specific recommendations concerning the minimal amount of training history (e.g. training logs) required to make probabilistically valid predictions for future performance. This is particularly critical in a military domain, where warfighter knowledge and skills must be stable and sufficient to succeed in any future maneuver or mission. First, we lay out basic tenets pertaining to this potential obstacle.

Resolution of Data Aggregate level data, by definition, reduces noise through averaging procedures that smooth out the shape of human performance curves. This process can be thought of as a double-edged sword. As a benefit, averaging helps reduce the contribution of noise to true human learning patterns. However, as a drawback, it is entirely possible that true human learning trends become masked or distorted as a result of the process (Estes, 2002). The magnitude of distortion could be caused by the amount of noise in the data, variability of parameter fits to individual trials or participants, and the range of variables of interest.

It may also be the case that producing an average group curve does not adequately *represent* the individuals it comprises, and further, the average group curve may not adequately *predict* individual performance. Chong and Wray (2005) provide evidence that the appearance of data at the aggregate level can be vastly different and even entirely distinct from curves using finer grains of analysis, so it is clear that these issues are not at all trivial at a practical level of utility.

An extensive literature review by Newell and Rosenbloom (1981) revealed that mathematically, learning trajectories of practice and retention at the aggregate level are generally best fit to power functions. Of interest is that learning trajectories at the individual level of performance are generally best fit to exponential functions (Heathcote, Brown, & Mewhort, 2000). This of course poses serious concerns for modeling purposes, as computational algorithms will always be best suited for data sets that have eliminated sources of spurious noise.

In order to make valid predictions or prescriptions of training regimen for individual warfighters, these tenets imply that it would behoove instructors to collect an adequate supply of data pertaining to training history, as data become more predictable when greater amounts of training history are initially utilized to baseline performance trajectories. This recommendation will become evident in the following sections.

Model Fits to Aggregate Level Data Using the CERI laboratory data, we initially tested model predictions at the aggregate level of performance, collapsing data across all individual team members and across all teams. In this evaluation scenario, we first optimized model parameters using performance history from the first day of testing. This required determining the values of learning and forgetting rates that best fit the performance function up to the end of day one training. As described above, the first day of testing required the completion of five, 40-minute reconnaissance missions, and is represented in Figure 5 as missions one through five.

After a 10-14 week delay, participants returned for a second session and engaged in missions six through eight. It is for these missions that we extrapolated mathematical regularities from known performance history to make our model predictions and compare against actual human performance. A correlation coefficient of 0.95 between the model and the humans was revealed, and is shown in Figure 5.

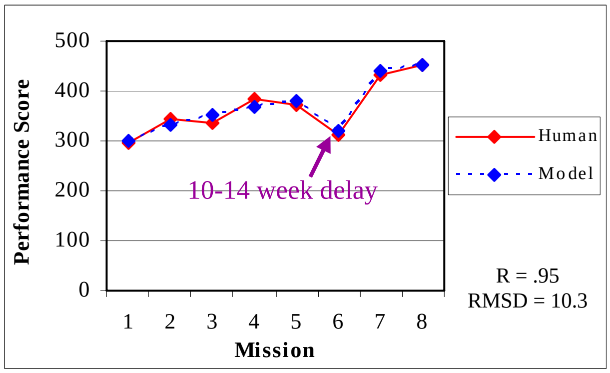

Figure 5: Predictive Performance Equation Model fit to aggregate level data after a 10-14 week delay

Model Fit to Individual Team Level Data Using the same procedure of optimization and extrapolation described above, we tested the efficacy of our model to make predictions at a finer grain of analysis, that being an individual team selected randomly from the sample. A correlation coefficient of 0.91 was revealed, producing the hypothesized reduction in predictive validity compared to the aggregate level, as shown in Figure 6.

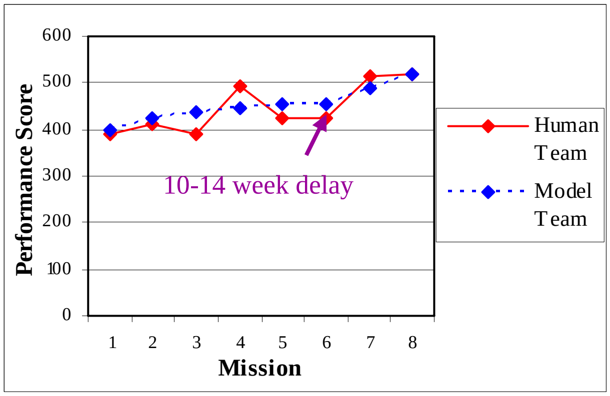

Figure 6: Predictive Performance Equation Model fit to team level data after a 10-14 week delay

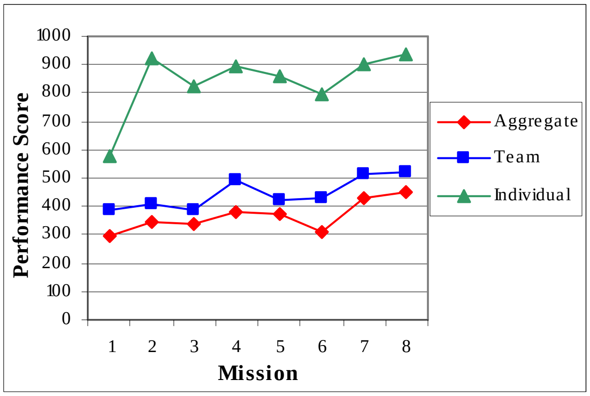

Figure 8: Comparison of true human performance curves at different levels of data resolution

Model Fit to Individual Operator Level Data Decomposing data down to the lowest grain of analysis in this data set, that of a randomly selected individual operator, further reduces the ability of the model to make accurate predictions. Increased noise in the data drops the correlation coefficient between the model and the human to 0.68, as shown in Figure 7.

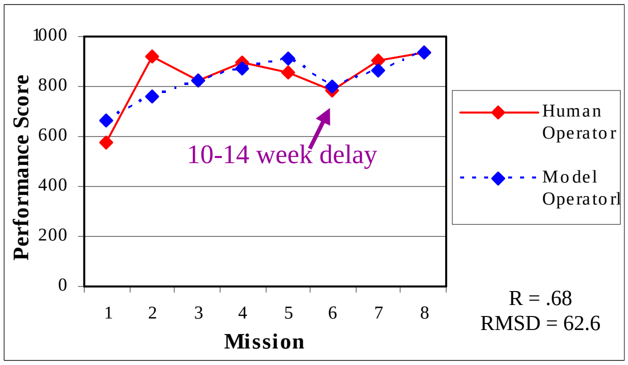

Figure 7: Predictive Performance Equation Model fit to individual operator level data after a 10-14 week delay

It is evident that performance curves at the individual team member, individual team, and overall aggregate levels can be very different and distinct from one another. Figure 8 illustrates this difference by presenting the randomly selected team and individual team member used in the model predictions, and compares them to the aggregate level performance curve.

This exercise also reveals how much poorer the prediction becomes when finer grains of analysis are used. More and more noise and error are introduced into the data when averaging procedures are removed; therefore, model predictions lose their mathematical base and fail in predictions of future performance. One useful way to combat this problem with finer grains of analysis would be to gather more information in training history, so that missions may be averaged across blocks for example, and noise and error would be systematically smoothed out.

Amount of Training History Another factor that affects model fits and future predictions is the amount of training history from which mathematical regularities are initially extracted. As such, we again used the CERI laboratory testbed data to examine model predictions dependent upon the amount of training history provided. For the previous predictions displayed at the aggregate, team level, and individual team member level performance, we optimized model parameters based on training from the first five missions (or session one of testing) to make predictions for the last three missions (or session two of testing, 10-14 weeks later). For this exercise, we compared model predictions as a function of the amount of training history at the aggregate level. We optimized model parameters from performance gleaned from one to seven known data points, and made predictions for the remainder of training. Not surprisingly, greater amounts of training history led to greater predictability in the data, and model efficacy rapidly increased with just four known points in training history. The correlation coefficients are displayed in Figure 9.

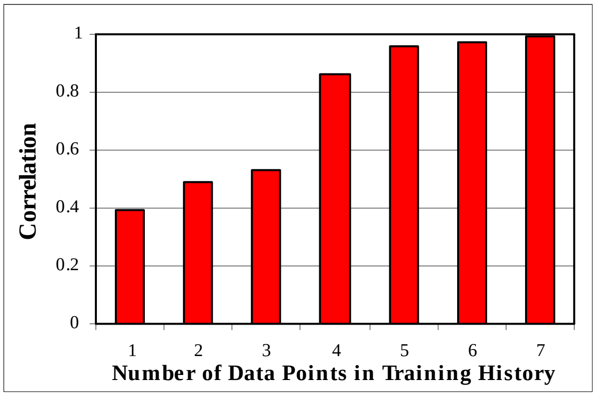

Figure 9: Predictive Model correlations to human performance data as a function of known training history

Clearly, this exercise of model predictability across varying amounts of performance history reveals the importance of collecting adequate amounts of performance data from the start. Stable learning trajectories allow the extraction of mathematical regularities to be implemented in a computational model, so that even at finer grains of analysis, the model may be useful in a predictive capacity.

Potential Predictive Utility in the Warfighter Domain

Of critical importance to the military and to individual warfighters themselves, is knowing when they have received enough training to be able to perform with consistency and to achieve success in specific missions or maneuvers at future points in time. Our predictive model has the potential to predict when a warfighter will achieve mission-readiness under very specific regimens of practice, with very specific distributions of practice. Take for example the following scenario: How long will it take an individual warfighter, using known performance training history, to achieve 95% proficiency under the current regimen of practice?

We constructed a hypothetical training scenario, based upon the design of the CERI laboratory study described above, to help illustrate the potential utility of our model. In this scenario, five 40-minute missions were completed in session one of training, an additional three 40-minute missions were completed in a second session between weeks 10 and 14 later, and our predictions for 95% proficiency at a later date were then extrapolated from the performance history of the first eight missions in total. Timetables for predictions were based on learners engaging in five missions per day at a rate of five days of training per week. Model results are presented in Figure 10, where performance history baseline is shown in blue, and model predictions are shown in red.

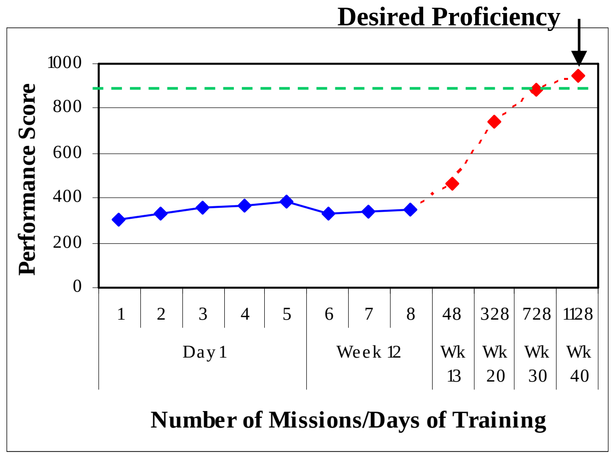

Figure 10: Notional prediction scenario

In this hypothetical example, the learner would require practice of an additional 1,120 40-minute training missions to achieve the desired level of proficiency. This translates to an additional 28 weeks of training above and beyond the baseline training period presented in blue, at a rate of five missions a day, five times a week.

This model is also equipped with the ability to make predictions for future performance using different specified regimens of practice, spaced apart at any length of time. Thus, if a learner takes two months away from training for instance, the model would be able to estimate how much knowledge had decayed over that period of time and make predictions for how much additional training would be required to achieve proficiency. This model therefore, has the potential to be a valuable predictive tool, even when training regimen is inconsistently spaced temporally or when extended breaks are taken.

Potential Prescriptive Utility in the Warfighter Domain

Also of great interest to the military, educators, and learners alike, is the development of a tool with the ability to prescribe optimal training regimens and maximize learning and retention gains. Our modeling tool has a potential prescriptive ability to assess and compare training schedules so that knowledge and skill acquisition will be more effective, and memory traces will be more durable over time.

Tapping into the history of empirical findings in the domain of learning and memory, it is clear that practices spaced further apart result in better retention than those spaced closer together, so this modeling tool may be used to predict and assess how effective each training repetition will be (as a function of memory trace activation) and to help optimize the spacing of training opportunities to result in larger learning gains.

Our predictive model carries the potential to function in these kinds of prescriptive capacities by means of hypothetical comparisons across learning opportunities spaced at varying points in time. Logistically, it can also help determine whether or not training expectations for achieving proficiency are feasible to accomplish within the specified boundaries of time; and if they are not, it may help inform trainers and educators as to what a more reasonable timetable would be. Take for example the following situation: How much training must an individual warfighter receive to be mission-ready (95% proficiency) by a specified deployment date four weeks away? Four months away?

We constructed a hypothetical training scenario, based upon the training design of the CERI laboratory study described in the preceding example, to help illustrate the potential prescriptive utility of our model. Again, we baselined the model parameters from the first eight missions of training and made predictions for the amount of training required to achieve 95% proficiency by each deployment date.

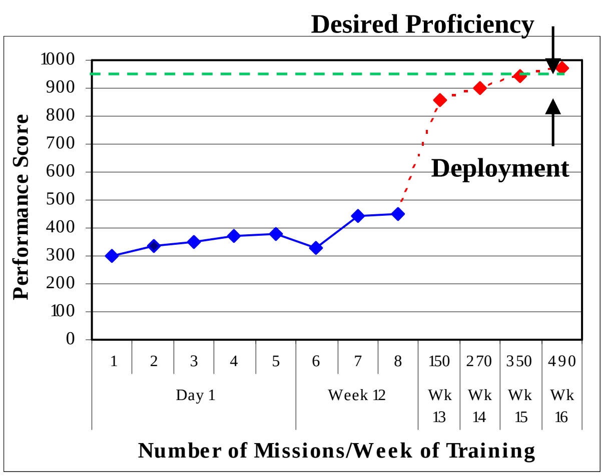

Figure 11: Notional prescription scenario – deployment date four weeks away

For the deployment scenario set four *weeks* away, this hypothetical warfighter would require approximately 120 40-minute practice missions to be completed each of the four weeks, to achieve mission-readiness (95% proficient) by that deadline (see Figure 11). This is of course an entirely unreasonable training expectation since it would require 24 training missions to be completed each day, and would barely allow any time at all for sleeping or eating. However, this is useful information, since the model may help point out when deployment dates are too early for warfighters to attain high enough levels of proficiency or to achieve high enough degrees of success. If there is no flexibility in deployment dates, this model may provide a reality check regarding expectations for readiness at the beginning of the deployment.

For the deployment scenario set four *months* away, this hypothetical warfighter would now require a more reasonable (but still aggressive) training regimen. The model calls for approximately 110 40-minute practice missions to be completed each of the four months, to achieve mission-readiness (95% proficient) by that deadline (see Figure 12). That's approximately five training missions each day, five days each week - a far more reasonable expectation than in the previous scenario.

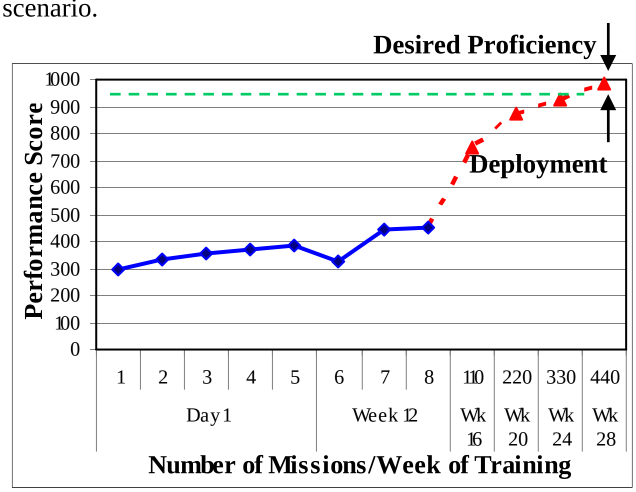

Figure 12: Notional prescription scenario – deployment date four months away

Also of interest with these deployment scenarios is the fact that training spaced further apart requires less overall training for the learner to actually achieve proficiency. There is a forty mission difference between the scenarios because learning gains are greater when training is distributed rather than massed. This fits nicely with well-established empirical data of human performance and shows the utility of the model for prescriptive and comparative purposes.

CONCLUSIONS AND FUTURE DIRECTIONS

We are enthusiastic regarding the potential uses for this type of model, particularly in the military domain. Use of this type of model can not only help determine when a warfighter has become proficient in a skill, but can also help streamline training to optimize learning as a whole. As these are initial tests of the model, additional analyses must be completed to further refine and validate the model. However, we are encouraged by the preliminary results and are hopeful we will have the opportunity to further investigate the model's strengths, limitations, and eventual uses.

ACKNOWLEDGMENTS

The authors would like to thank the Cognitive Engineering Research Institute (CERI), and particularly Dr. Nancy Cooke, for providing data for our model's evaluation. A portion of this research was completed during the first author's internship at the Air Force Research Laboratory, facilitated by an appointment with the Oak Ridge Institute for Science and Education.

REFERENCES

Aleven, V., & Koedinger, K. R., (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science, 26(2)*, 147-179.

Anderson, J. R. (1983). A general learning theory and its application to the acquisition of proof skills in geometry. In R. Michalski, J. Carbonell, and T. Mitchell (Eds.), *Machine Learning: An Artificial Intelligence Approach*. Palo Alto, CA: Tioga Publishing.

Anderson, J. R. (1992). Automaticity and the ACT* theory. *American Journal of Psychology*, 105, 165-180.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y . (2004). An integrated theory of the mind. *Psychological Review 111*, (4). 1036-1060.

Anderson, J. R., Conrad, F. G., & Corbett, A. T. (1989). Skill acquisition and the LISP Tutor. *Cognitive Science, 13*, 467-506.

Anderson, J. R., & Fincham, J. M. (1994). Acquisition of procedural skills from examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 239-277.

Anderson, J. R., Fincham, J. M. & Douglass, S. (1999). Practice and retention: A unifying analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 1120-1136.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Anderson, J. R., & Schunn, C. D. (2000). Implications of the ACT-R learning theory: No magic bullets. In R. Glaser (Ed.), *Advances in instructional psychology: Educational design and cognitive science, Vol. 5*. Mahwah, NJ: Erlbaum.

Bahrick, H., & Phelps, E. (1987). Retention of spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 344—349.

Barnes, C. A. (1979). Memory deficits associated with senescence: A neurophysiological and behavioral study in the rat. *Journal of Comparative Physiology, 43*, 74-104.

Bean, C. H. (1912). The curve of forgetting. *Archives of Psychology, 2*, 1-47.

Chong, R., & Wray, R. (2005). Inheriting constraint in hybrid cognitive architectures: Applying the EASE architecture to performance and learning in a simplified air traffic control task. In K. Gluck and R. Pew (Eds.), *Modeling Human Behavior with Integrated Cognitive Architectures*.

Ebbinghaus, H. (1885). *Memory: A Contribution to Experimental Psychology*. Trans. H.A. Ruger and C.E. Bussenius, New York: Columbia University Press.

Estes, W. K. (2002). Traps in the route to models of memory and decision. *Psychonomic Bulletin & Review, 9(1)*, 3-25.

Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning & Verbal Behavior, 15,* 1-16.

Greer, J. E., & McCalla, G. I. (Eds.) (1994). *Student modeling: The key to individualized knowledge-based instruction*. New York: Springer-Verlag

Hartley, J. R., & Sleeman, D. H. (1973). Towards more intelligent teaching systems. *International Journal of Man-Machine Studies, 5*, 215-236.

Heathcote, A., Brown, S., & Mewhort, D. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin and Review, 9*, 394-401.

Koedinger, K.R., & Anderson, J.R. (1993). Reifying implicit planning in geometry. Guidelines for model-based intelligent tutoring system design. In Lajoie, S., & Derry, S. (Eds.) *Computers as Cognitive Tools*. Hillsdale, NJ: Erlbaum.

Neves, D. M. & Anderson, J. R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition*. Hillsdale, NJ: Erlbaum.

Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive Skills and their Acquisition.* Erlbaum: Hillsdale, NJ.

Pavlik, P. I, & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation based model of the spacing effect. *Cognitive Science, 29*, 559-586.
Shute, V. J., & Psotka, J. Intelligent tutoring systems: Past, present, and future. In D. H. Jonassen (Ed.) *Handbook of Research for Educational Communications and Technology*, 570-600. New York: Macmillan.
Strong, E. K. (1913). The effect of time interval upon recognition memory. *Psychological Review, 30*, 339-342.
VanLehn, K. (1988). Student modeling. In M. C. Polson & J. J. Richardson (Eds.) *Foundations of Intelligent Tutoring Systems*. Mahwah, NJ: Lawrence Erlbaum Associates.
Woodworth, R. S. (1938). *Experimental Psychology*. New York: Henry Holt and Company.