Oury, J. D., Tehranchi, F., & Ritter, F. E. (2018). Predicting learning and retention of a complex task. In Proceedings of the 16th International Conference on Cognitive Modeling (ICCM 2018). 90-95.

Predicting Learning and Retention of a Complex Task

[1]Jacob D. Oury (OuryJacob@psu.edu)
[2]Farnaz Tehranchi (farnaz.tehranchi@psu.edu)
[1]Frank E. Ritter (frank.ritter@psu.edu)
[1]College of Information Sciences and Technology, Penn State, University Park, PA 16802 USA
[2]Department of Computer Science and Engineering Penn State, University Park, PA 16802 USA

Abstract

We use an ACT-R model of a complex task to explore the implications of ACT-R's learning and forgetting mechanisms to better understand learning and retention. The model performs a task that has 14 non-iterated subtasks that takes approximately 25 min. to perform the first time. The results show that a typical learning curve is generated by the model that is well fit to human data. When decay is examined we find that the retention curves basically match the shapes predicted by the KRK theory, and that training and testing have been confounded in many studies. From these results we see that the previously hypothesized mixed declarative procedural stage of learning actually starts on the first trial and is never completely exited, so we will need to propose other thresholds to mark transitions between declarative, mixed, and proceduralized knowledge. We predict based on this model that learning and retention will vary greatly by task components, practice schedule, and learner's strategy.

Keywords: ACT-R; learning curve; retention model

Introduction

Kim, Ritter, and Koubek (2013) provided a summary of learning theories in a review paper. Their summary theory is based on learning theories by Fitts (1964), Anderson (1982), Rasmussen (1986), and VanLehn (1996). It is also consistent with further work reviewed by Kim et al. (2013), as well as other theories of learning (e.g., Posner, 1973) and data on learning (e.g., Seibel, 1963)–the diagrammatic representation of this theory is shown in Figure 1.

In this paper we first briefly review the work this theory is based upon to suggest how to better test it. We then use the ACT-R cognitive architecture to make predictions about learning and forgetting, including where the stages might appear. We then test the predictions of the theory using data (Kim & Ritter, 2015), and use the model to design a further empirical study to test and illustrate the theory's predictions.

The KRK Theory

This review starts with the KRK theory and its predictions. Some of the data and theories that it was based on and some further work, both empirical and theoretical is examined to find further support and limitations. Implications for further development are provided in a summary.

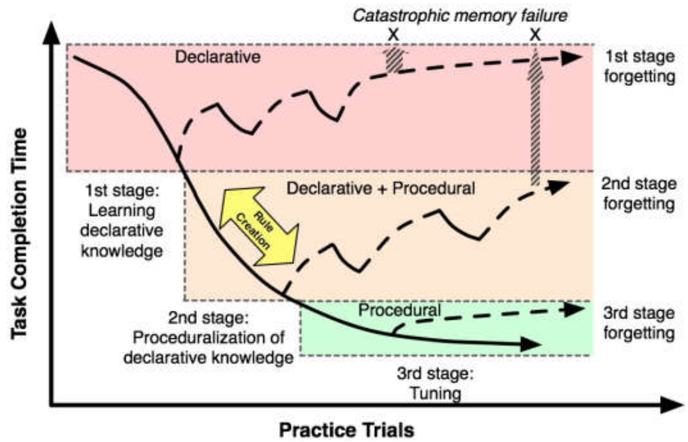

Figure 1. The KRK theory of learning. This represents a summary theory of predictions 1-5 from Table 1. Taken from Kim, Ritter, and Koubek (2013).

The KRK theory was developed as part of Kim's PhD thesis (Kim, 2008). The theory, shown in Figure 1, makes several predictions implicitly. Table 1 explicitly shows predictions from Figure 1 and new predictions. In a larger paper we explain these predictions in more detail (Ritter et al., forthcoming). Here, we simply summarize them before testing them.

Of the nine hypotheses in Table 1, items 1-6 are basically supported in the Kim et al. (2013) review paper. Items 7-9 are new and come from the inclusion of perceptual-motor and recognition memory. The empirical support for the hypotheses are piece-wise and often on simple tasks (e.g., Choice Reaction Theory, Seibel, 1963). The forgetting curves often are from single points of learning (e.g., Kim 2008). That is, we do not know of a single study that predicts all these curves, and most of the studies that are used to derive and support these hypotheses use simple tasks, such as word association.

It would be useful to explore these predictions with a complex, multi-step task and to explore the whole set of predictions with a single empirical study with longer retention intervals. This study would be a large undertaking. So we will use a model of a complex task in an architecture with learning and forgetting that has multiple skill representations to explore the study first and see what the model's predictions are for learning and retention.

Table 1. Human performance hypotheses from the KRK theory. Items 1-6 are supported by Kim et al. (2013). 7-9 are new predictions from incorporating new memory types.

Prior Predictions	
(1)	Learning follows the power law curve of learning Time = A + BN^{-C} (A, B, C are constants)
(2)	Three stages of knowledge: a. Acquiring declarative and procedural knowledge b. Consolidating the acquired knowledge c. Tuning the knowledge towards overlearning
(3)	Retention of declarative knowledge decays quickly and catastrophically
(4)	Retention of mixed declarative and procedural knowledge decays moderately
(5)	Retention of proceduralized knowledge has least decay
(6)	Recognition and perceptual-motor knowledge have different learning curves than procedural or declarative.
New Predictions	
(7)	Ideal training schedules will vary by knowledge-type; perceptual-motor may require minimum training block size
(8)	Retention of perceptual-motor knowledge appears to decay little
(9)	Recognition memory is (probably) not fragile

Method

We will first describe a complex task that we use and for which we have some data and a running model that learns. We will then describe the architecture and the model, which serves as a subject in this simulated study. We then describe the human data, and how we ran the model.

The Dismal task

The Dismal task, illustrated in Figure 2, is a spreadsheet task that can be used to measure procedural knowledge and skills learning and decay (Kim & Ritter, 2015). The Dismal task was created to be done in the Dismal spreadsheet (Ritter & Wood, 2005). Dismal is an open source, extendable spreadsheet in Emacs.

The overall task length and subtask variety make Dismal a relatively complex task. There are 14 different subtasks in Dismal (Table 2). The subtasks contained attention shifts, encoding of information, attending to information, key presses, and mouse moves/clicks. Previous work using Dismal allows comparison to human data and model predictions.

These tasks can be done with two different interfaces—(a) a keyboard with key-based commands, or (b) a mouse or vertical mouse. The vertical mouse provides new motor skills to learn and forget because it requires a different hand posture.

Table 2. The 14 Dismal spreadsheet subtasks.

Dismal task sequence	
(1)	Open a file, named normalization.dis under the “experiment” folder
(2)	Save as the file with your initials
(3)	Calculate and fill in the frequency column (B6 to B10)
(4)	Calculate the total frequency in B13
(5)	Calculate and fill in the normalization column (C1 to C5)
(6)	Calculate the total normalization in C13
(7)	Calculate the length column (D1 to D10)
(8)	Calculate the total of the “Length” column in D13
(9)	Calculate the Typed Characters column (E1 to E10)
(10)	Calculate the total of the “Typed Characters” column in E13
(11)	Insert two rows at A0 cell
(12)	Type in your name in A0
(13)	Fill in the current date in A1 using the command
(14)	Save your work as a printable format

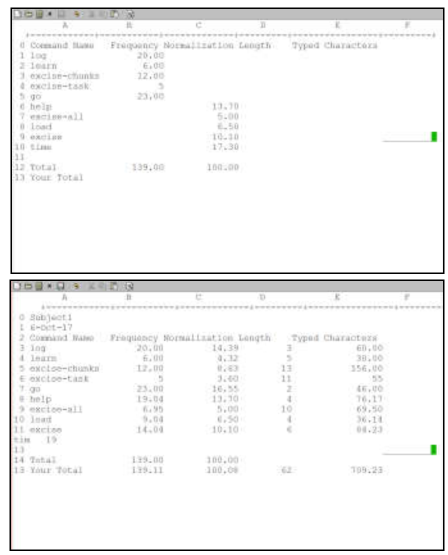

Figure 2. Dismal interface with initial state (top) and final state of the task (bottom).

The two interface modes can be used to study different types of knowledge: recall of keystroke commands and recognition of menu-based commands.

ACT-R

ACT-R is a theory of the mechanisms that make up cognition. It is an example of a unified theory of cognition (Byrne, 2012; Newell, 1990), in that intends to predict and explain human behaviour by simulating the steps of cognition with a fixed set of mechanisms. ACT-R predicts behaviour and activation of brain regions by using mechanisms including procedural and declarative knowledge, and working memory as activation, to perform tasks.

We briefly review ACT-R’s components and then the memory equations; other more complete treatments are available (e.g., Anderson 2007; Anderson 1982). Thus, we briefly review ACT-R’s components and then the memory equations.

The architecture components

ACT-R consists of modules and buffers. Modules are responsible for processing one kind of information and are the mechanisms for modifying and implementing a buffer; buffers are contents that are visible to other modules. The modules descriptions and roles include:

the Visual module is for identifying objects in the visual field. Visual objects and their identities are located in the visual buffer and visual location buffer and monitoring attention and visual objects such as scanning a computer screen.

the Manual module is for controlling the hands; the manual/motor buffer handles controlling and monitoring hand movement such as typing on a keyboard.

the Declarative module is for retrieving information from memory, and a Goal module is for keeping track of current goals and intentions. The Goal buffer stores the current subgoal step and its next step.

A central production system is a rule-based system that performs the matching, selection, and execution of production rules. Also, it coordinates the communication and performance of these modules through the application of production rules. The central production system works in parallel with modules and constantly updates and queries the buffers' data.

The memory equations

Throughout the task completion by an ACT-R model, each declarative memory used will have its base-level activation increased. Presentation of an item (or chunk), and thus subsequent changes in base-level activation, can occur at three points in the process: at item/chunk creation, at the time when two items are merged, and when the chunk itself is retrieved.

This process has two mutually exclusive options for how to calculate learning for chunks during the task procedure: the Optimized Learning Equation (OL), and the Base-Level Learning Equation (BL),. These are shown in Equations 1 and 2. Parameters (e.g., :bll) refer to specific values set within ACT-R's base configuration.

Equation 1: The Optimized Learning Equation (OL)

$$B_i = \ln\left(\frac{n}{1-d}\right) - d * \ln(L) + \beta_i$$

***n*:** The number of presentations for chunk *i*
***d*:** The decay parameter set using the :bll parameter
β_i: A constant offset set using the :blc parameter

Equation 2: The Base-Level Learning Equation (BL)

$$B_i = \ln\left(\sum_{j=1}^{n} t_j^{-d} + \beta_i\right)$$

***L*:** The lifetime of chunk *i* (the time since its creation).
t_j: The time since the *jth* presentation. A presentation, or reference, is either the chunk's initial entry into DM or when another chunk is merged with a chunk
Unlabeled variables in BL are shown in Equation 1.

The two equations differ in their accuracy and computational cost. BL is costlier because it accounts for time through the t_j parameter (with associated repeated exponential computations based on t_j's), while OL simplifies the equation to primarily rely on the number of presentations.

The model

Herbal, a high-level behaviour representation language, creates ACT-R source code (Ritter et al., 2006). Herbal has been used to build several ACT-R models of this spreadsheet task that ranged in expertise from novice to expert. Further details about the mode can be found in Paik, Kim, Ritter, & Reitter (2015). We used the novice model with 9 initial rules and 520 declarative memory elements because it starts the task using declarative knowledge.

Number of runs

Completion time curves were generated from the mean time for a given data point based on multiple runs of the model for the task (N=5). This number is sufficient because we are observing broad trends rather than focusing on specific effect sizes (Ritter, Schoelles, Quigley, & Klein, 2011).

Existing human data

Participants in Kim's (2008; Kim & Ritter, 2015) study (N=60) were divided equally into two groups: one used only the keyboard (N=30), and another group used the combination of vertical mouse and keyboard (N=30). They completed the Dismal task and came back at 6, 12, or 18-day intervals. We used the vertical mouse interface subjects' data. The human data for days 1-4 matches the model predictions for trials 1-4 shown in Figure 3.

Results

We ran a series of models to explore how the ACT-R model of the Dismal task predicts learning and decay. For each test, the model was run with the OL and BL equations to compare their predictions. We found similarly shaped retention curves for both, but there were differences as well.

Predictions with Optimized Learning and Decay

Figure 3 shows a standard learning curve on task time where the model was run over 10 trials without delays between trials. Decay curves are shown for up to 5 days decay.

This is how most repeated trial ACT-R models are run. They are run multiple times, with no time between trials in the model, even if there is time between human trials for the subjects, which there was for this data set–each trial was run on a separate day.

These results are consistent with most of the predictions in Table 1. (1) Task completion times on the solid black line followed the shape predicted by the power law of learning. (2) The three stages are there, but when examining the model trace, we see that proceduralization starts in the first

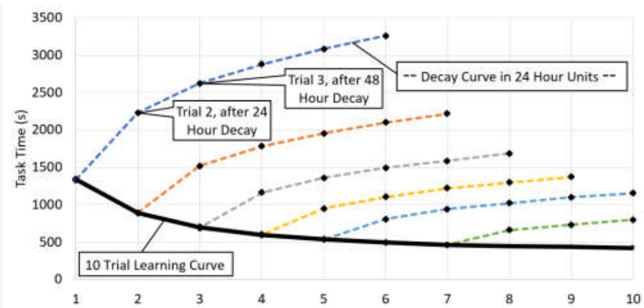

Figure 3: Predictions for task time with OL (black). Forgetting curves show task time after [1-5] days of decay after a period of consistent practice.

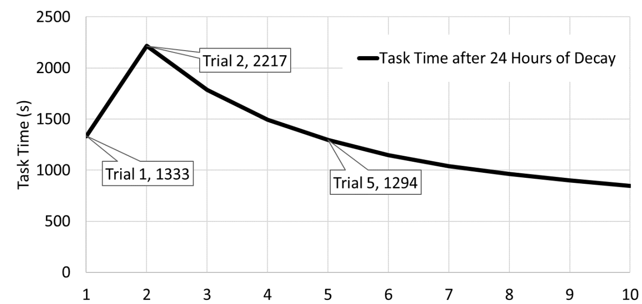

Figure 4: Predictions for task completion time with OL for 10 trials with 24-hour decay periods between each trial.

task of the first trial. (3) Declarative knowledge decays but did not do so catastrophically for this model and task. (4 & 5) the decay of the mixed knowledge (middle trials) was slower than the declarative knowledge but not as slow as the procedural knowledge (later trials). We could not examine the other predictions (7,8,9) with this task.

We found unexpected results from applying a decay period in the model. After one practice and 24 hours of decay, the task time was slower than during the first trial. Decay caused a worse time than a novice's first trial. The mean task time for trial 1 was 1335 s. After the 24-hour decay period, task time for trial 2 was 2228 s compared to 882 s for the normal curve. In short, decay causes memories (and thus performance) to be worse than the very first trial of the experiment.

So, we explored the effects of decay by including a 24-hour decay period after each practice to simulate what the subjects did, that is, practice and then wait a day until the next practice. These results are shown in Figure 4.

Figure 4 shows a typical learning curve after trial 2. The initial 24-hour decay causes performance to be longer on day 2 than on the first day. The performance does not get as fast as day 1 until trial 5. Including the time between practices does not improve our predictions. We do not include forgetting curves because the learning curve is unusual.

Predictions with base-level learning equation

Next, we used the base-level learning equation for the same process. Figure 5 shows the decay and retention curves through 10 trials of practice with the base-level learning

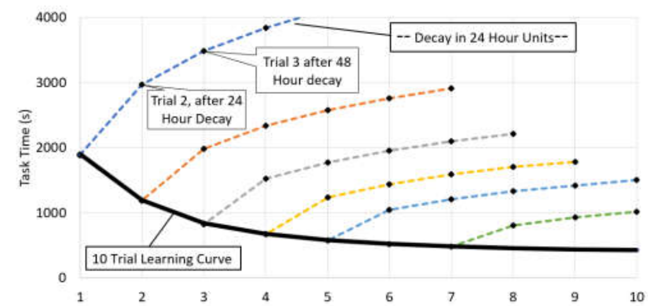

Figure 5: Predictions for task time with BL for ten trials of practice (black). Forgetting curves show task time after [1-5] days of decay after a period of consistent practice.

Figure 6: Predictions for task completion time with BL for 10 trials with 24-hour decay periods between each trial.

equation instead of the optimized learning equation. Figure 5 is without time between trials and Figure 6 is with 24 hours between trials. In Figure 6 we do not include forgetting curves to emphasize the unusual shape of the learning curve.

Overall, the learning and retention curves with BL are shaped similarly to the OL model predictions but start with a higher mean task time at trial 1 (1901 s). The final task time was 424 s, or 22% of the first trial. Again, this showed that the performance on trial 2 following 24-hour decay was worse than the initial time.

Predictions with initial day delay and day delays inserted

We then considered applying a day delay in the initial declarative knowledge before starting to perform the task. This could represent less complete declarative learning of the task knowledge. So, we set the model to run with a day delay after the declarative knowledge has been first learned.

Figure 7 (top) shows three learning curves using the OL equation, and the BL equation (bottom). The red dashed curve is the learning curve without time decays; it is the same as Figure 3. The blue dotted line is with 24 hours between each trial, the same as Figure 4. The black solid line has a 24-hour decay before the first trial and after each trial.

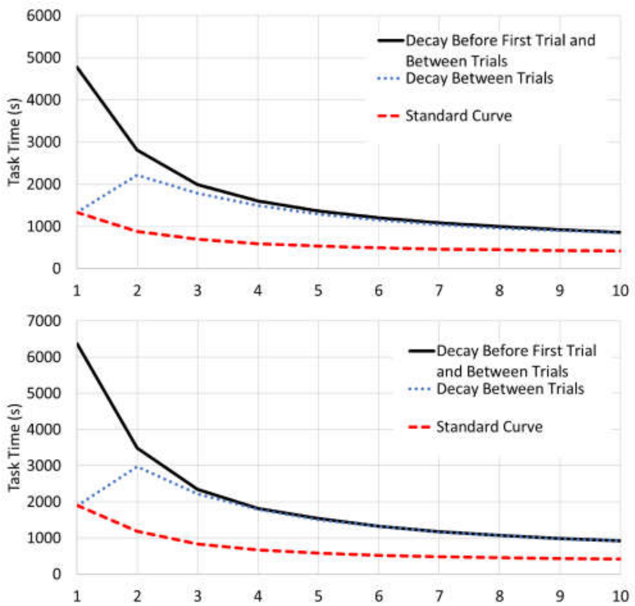

Figure 7: Predictions for task completion time with OL (top) and BL (bottom) with 24-hour decay periods between each trial and an initial 24 decay on the declarative knowledge.

Final model, base-level learning with day delays and retention curves

The predictions in Figure 7 remain somewhat unsatisfactory. They either leave out the decay between practice trials (but fit the data fairly well), or they include the decay, but over predict the task time. We thus tried an adjustment to the decay parameter suggested by Lebiere (personal communication, October 2017).

This adjustment provides different decay constants for within and without an experimental situation. This change probably represents the effect of proactive interference more accurately in that during a study the task-related memories are more similar in a time block in a study than they are outside a study. With non-study time, the decay time is reduced to be ¼ of the actual delay time, that is, instead of 24 h between trials, only 6 hours is added as decay after learning the task knowledge and between trials. Figure 8 shows this model's predictions.

Discussion and Conclusions

This model provides several insights about learning and the ACT-R memory equations.

Insight: There are no fixed learning stages

The results of the model's learning show that there are no crisp divisions between the learning stages, contrary to predictions made in previous theories for the learning curve – there are no points of inflection in the learning curve to show the transition. In the knowledge used in the model (declarative and procedural), there are also not inflection points. The model thus predicts that there are not distinct stages in this task at least, where the knowledge is

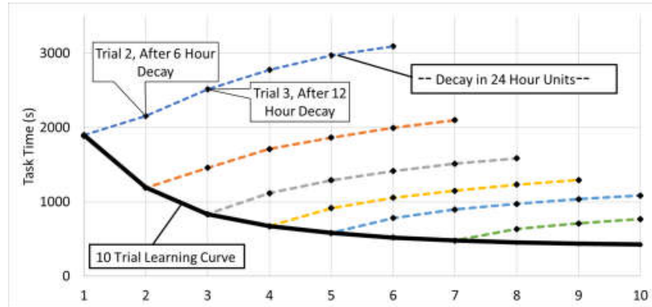

Figure 8. Predictions for task completion time with BL with 24-hour decay periods between each trial and an initial 24 decay on the declarative knowledge, adjusted

all declarative or all procedural. Subtasks are more clearly one or the other because only one type of knowledge might be used in a smaller task. However, the model shows that even during the first trial, knowledge is being not only proceduralized, but also that the procedural knowledge is being used. So, partway through the first trial, the model has mixed task knowledge in some technical sense.

We suggest that the stages be relabeled into mostly declarative, mixed, and nearly all procedural. This is in contrast to [all] declarative, mixed, and [all] procedural.

Insight: the stages will vary by task and strategy

These stages will also depend on the task components and the task distribution. If the task is primarily a declarative task, it will basically stay a declarative task. If it is a perceptual-motor task or a procedural task, it will transfer into a proceduralized task.

This model also shows that location of these stages will vary by task. Tasks with high declarative components will remain in the mostly declarative stage longer, according to this definition, because more parts that will stay declarative.

If an unusual task from the distribution of possible tasks comes along, the learner may be shifted back towards more declarative task knowledge, or as Rasmussen (1983) notes, knowledge-based control knowledge. So, distribution of tasks that all occur equally will have different learning and stages than a distribution of same tasks (and knowledge) where some only occur rarely.

Insight: implications for empirical studies

The model provides some implications for running a human study in this area. To measure the decay curve, you must have at least three decay points or two points and a strong theory. A study with humans cannot reset the model to get multiple decay measurements but will have to train a subject and then can only measure decay once without retraining occurring. To measure these points, you must train subjects to standard, and then have them come back at a delay. You can only have them come back once, because the measure is a training. Thus, to measure the decay at three points, you have to each data point be its own condition. To study decay after training 1, 2, and 5 days, and decay at 4, 8, and 16 days, you require nine groups. Thus, the decay graph 9x more expensive (assuming subject drop out does not increase because of the delay) than the simple learning curve. This is partly why these curves are studied less.

Insight: The decay curve at one day is practice on the next day

Consider that you will be training every day and wish to study the decay function after 1, 2, and 5 consecutive practice days. If you measure the amount of decay after one day, the performance after one day of decay when you are training once per day is the same as the group that is on the training schedule. Thus, the decay curve in such circumstance has a decrease in performance time with one day decay (if the test includes training such as doing the task), and then decreased performance on later days. This may be like walking as being way of falling forward.

Insight: The subtask curves within most instruction are a mix and have varied retention intervals

Very few real-world tasks will have this pure of a training and retention schedule. In the real world, after an hour of training, the learner will move into new material in later sessions. Thus, the learning and retention curves will include multiple small curves, and some subtasks will be trained every session and get much faster, and some tasks will occur only rarely and will have long decay times (if learned early) or short decay times (if learned later). You would need a computer to keep track of them!

Future Work: How to run a study of a complex task to test the KRK theory

Based on these results, in our empirical test of the KRK theory we will examine a complex task, a trouble shooting task, with 1, 2, and 5 practice trials, separated by a day per trial. We will look at performance at 3, 5, and 7 days decay after the last practice. We will not have the full training material at the decay tests, just the trouble shooting tests. This will measure the decayed knowledge with little or no relearning of it.

Acknowledgments

Kevin Gluck gave useful comments about this model in a discussion at ICCM. This project was supported by ONR grant N00014-15-1-2275.

References

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369-406.

Byrne, M. D. (2012). Unified theories of cognition. *Wiley Interdisciplinary Reviews: Cognitive Science, 3*(4), 431-438.

Destefano, M. (2011). *The mechanics of multitasking: The choreography of perception, action, and cognition over 7.05 orders of magnitude.* Unpublished PhD thesis, Cognitive Science, RPI.

Fitts, P. M. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), *Categories of human learning* (Vol. 47, pp. 381-391). New York: Academic Press.

Kim, J. W. (2008). *Procedural skills: From learning to forgetting.* Unpublished PhD thesis, Department of Industrial and Manufacturing Engineering, The Pennsylvania State University, University Park, PA.

Kim, J. W., & Ritter, F. E. (2015). Learning, forgetting, and relearning for keystroke- and mouse-driven tasks: Relearning is important. *Human-Computer Interaction, 30*(1), 1-33.

Kim, J. W., Ritter, F. E., & Koubek, R. J. (2013). An integrated theory for improved skill acquisition and retention in the three stages of learning. *Theoretical Issues in Ergonomics Science, 14*(1), 22-37.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Paik, J., Kim, J. W., Ritter, F. E., & Reitter, D. (2015). Predicting user performance and learning in human-computer interaction with the Herbal compiler. *ACM Transactions on Computer-Human Interaction, 22*(5), Article No.: 25.

Posner, M. I. (1973). *Cognition: An introduction*. Glenview, IL: Scott Foresman.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-13*, 257-266.

Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering.* New York, NY: Elsevier.

Ritter, F. E., Haynes, S. R., Cohen, M. A., Howes, A., John, B., Best, B., et al. (2006). High-level behavior representation languages revisited. In *Proceedings of ICCM - 2006- Seventh International Conference on Cognitive Modeling*, 404-407. Edizioni Goliardiche: Trieste, Italy.

Ritter, F. E., Schoelles, M. J., Quigley, K. S., & Klein, L. C. (2011). Determining the number of model runs: Treating cognitive models as theories by not sampling their behavior. In L. Rothrock & S. Narayanan (Eds.), *Human-in-the-loop simulations: Methods and practice* (pp. 97-116). London: Springer-Verlag.

Ritter, F. E., & Wood, A. B. (2005). Dismal: A spreadsheet for sequential data analysis and HCI experimentation. *Behavior Research Methods, 37*(1), 71-81.

Seibel, R. (1963). Discrimination reaction time for a 1,023-alternative task. *Journal of Experimental Psychology, 66*(3), 215-226.

Tehrachi, F., Oury, J.D., Ritter, F. E., Predicting learning and retention (forthcoming).

Tenison, C., & Anderson, J. R. (2016). Modeling the distinct phases of skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(5), 749.

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology, 47*(513–539.