In D. Reitter & F. E. Ritter (Eds.), Proceedings of the 14th International Conference on Cognitive Modeling (ICCM 2016). University Park, PA: Penn State.

Microgenetic Analysis of Learning a Task: Its Implications to Cognitive Modeling

Jong W. Kim (jong.kim@ucf.edu)
University of Central Florida
Orlando, FL 32816 USA

Frank E. Ritter (frank.ritter@psu.edu)
College of Information Sciences and Technology, Pennsylvania State University
University Park, PA 16802 USA

Abstract

We report a microgenetic and quantitative analysis of a large learning data set. We analyzed performance change by four practice trials (once per day) and by the 14 different subtasks with more than 500 total keystrokes. Specifically, we compared performance change across the subtasks—some subtasks are cognitive problem-solving and others are perceptual-motor driven tasks. This microgenetic approach provides an understanding of how a local performance in a task affects the global performance of a whole task. We computed the learning curve constants for the different subtasks. We found evidence to support the KRK theory of learning and retention (Kim & Ritter, 2015). The results suggest that learning varies by subtask and by its type.

Keywords: Microgenetic analysis; Learning; Cognitive modeling.

Introduction

In general, learning can be described as a speed-up or practice effect (Ritter, Baxter, Kim, & Srinivasmurthy, 2013; Seibel, 1963). To help better understand our learning performance, it is necessary to focus on a couple of variable factors in tasks and their types. Complex tasks may consist of different components of subtask skills. Presumably, different subtask skills may be learned and retained in our memory. This understanding would affect the perspectives of learning, learning environments, instructional systems (e.g., contents), and interface design of such systems.

As one small step contributing to learning research, we investigated learning and retention of a complex task consisting of the 14 subtasks by comparing two input modalities (Kim & Ritter, 2015). This investigation suggests that the prevalence of GUI interfaces can be attributed to a more relearnable design compared with a keystroke-based interface, and suggests more investigation on where learning (and forgetting) occur during the course of complex tasks.

In this paper, we conduct a deeper analysis; a microgenetic analysis of learning in an attempt to identify how learning is different across the 14 subtasks. We look at individual subtask skill components over four practice trials. This approach is similar to a microgenetic study examining sources of change in cognitive development and learning (e.g., Agre & Shrager, 1990; Moon & Fu, 2008; Siegler, 2006). It is expected that our approach can provide a deeper understanding of where learning occurs and how different knowledge types are learned.

Learning as a Whole Task

A considerable amount of literature suggests a consensus understanding of learning; a three-stage process of learning provides a theoretic account of performance change including (a) acquiring declarative knowledge from instruction to perform a task in the first stage, (b) consolidating the acquired knowledge into a procedural form with practice in the second stage, and (c) tuning the knowledge toward overlearning exhibiting the speedup effect of the knowledge application mental process (Anderson, 1982; Fitts, 1964; Rasmussen, 1986). Based on this consensus foundation of learning, a study of forgetting expands how an individual learns and retains knowledge and skills theoretically, empirically, and computationally (Kim & Ritter, 2015), shown in Figure 1. A widely used cognitive architecture, ACT-R, implements the computational features of the three-stage process by proposing that performance change follows a regularity known as the power law of practice—the time to complete a task speeds up with practice according to a power function (e.g., Anderson, Fincham, & Douglass, 1999; Newell & Rosenbloom, 1981; Seibel, 1963). An exponential function is also widely accepted to summarize the practice effect (e.g., Heathcote, Brown, & Mewhort, 2000).

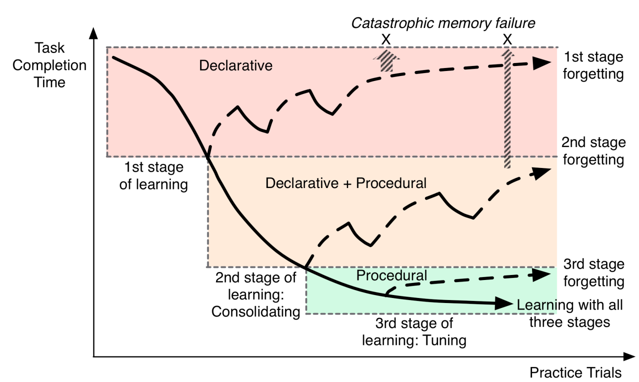

Figure 1: The KRK three-stage learning and retention theory (Kim & Ritter, 2015).

Recent reports provide a predictive analysis of a spreadsheet task, the Dismal spreadsheet task (Kim & Ritter, 2015; Paik, Kim, Ritter, & Reitter, 2015), using KLM-GOMS (Card, Moran, & Newell, 1983) and ACT-R (Anderson, 2007). These analyses examine performance change from a novice through an intermediate to an expert performing a complex task. The task includes subtasks, and the time to complete the task is predicted by the aggregate resources subtasks use (i.e., cognitive, perceptual, and perceptual-motor skills). These predictions can be meaningfully decomposed to each subtask skill and single action, and these can be compared with the data on the same level, providing an organization for a microgenetic analysis.

A Baseline Prediction

As a baseline prediction, a KLM-GOMS model was used to predict error-free expert performance on the Dismal task. The task completion times were computed to be compared with ACT-R predictions and the human data of the whole task of the Dismal spreadsheet task. The model includes primitive physical-motor operators (*K* – keystroke, *P* – pointing, *H* – homing, and *D* – drawing), mental operators (*M*), and system response time (*R*), as shown in Equation (1). In the interest of simplicity and because of relatively fast response times, we ignored the system response time (T_R=0).

$$T_{execute} = T_K + T_P + T_H + T_D + T_M + T_R \qquad (1)$$

We used three physical-motor operators (*K*, *P*, and *H*) and the mental operator (*M*) for time predictions of the Dismal spreadsheet task. The default time was used for homing and mental operators. During the mental operator time (T_M), participants mentally prepare what to press and retrieve items from memory including the next step. We followed the existing heuristic rules for determining the use of mental preparation (Card, Moran, & Newell, 1983, p. 265) and used the default time, 1.35 s. We placed a mental operator in front of all pointing activities (pointing to a menu item) and all key-press activities (pressing a keystroke command). To complete the first subtask (Open File), theoretically, participants in the keyboard group needed 3 mental operators (refer to Table 2). The homing time (T_H) for hand movements between different physical devices was 0.4 s.

To calculate the keystroke time (T_K), we know that it varies across individuals. We, therefore, computed the time from the first keystroke to the last in the first subtask for both modalities. The average keystroke time ranged from 0.95 s/keystroke on the first day of learning to 0.47 s/keystroke on the last day of learning. If we refer to the keystroke time in Card et al. (1983, p. 264), our data indicate the participants' keystroke speed resided between the worst typist, 1.20 s and the speed of average non-secretary typist, 0.28 s. We used 0.47 s for the T_K parameter as an expert performance. Shift and control keys were counted as a separate keystroke. The predicted task completion time for users in the keyboard group was 666.67 s as seen in Table 2. We present the details of the KLM-GOMS analysis of each subtask in the Microgenetic Analysis of Learning section.

ACT-R Prediction

Several cognitive architectures predict learning, which is beyond the capability of KLM. Particularly, the ACT-R architecture provides predictions of performance changes due to learning. Furthermore, the ACT-R model can predict learning on this task from a novice to an expert, as shown in Figure 2. The model consists of production rules and declarative memory elements to represent practice effects, which can be compared with human learning data (Paik et al., 2015).

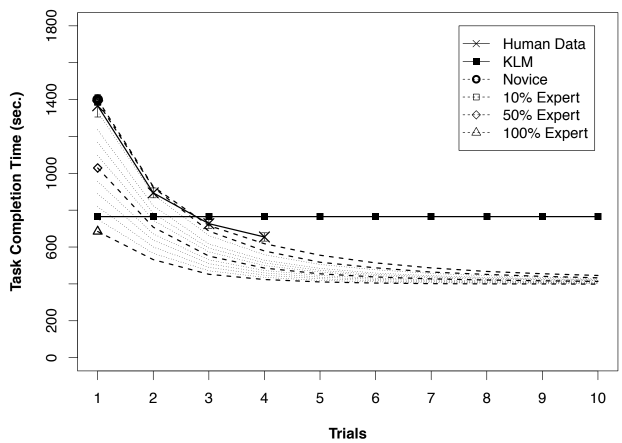

Figure 2: ACT-R models of the Dismal task (dashed lines, from fully novice to previously practiced expert), along with human aggregate data (X's and SEM error bars), and the KLM prediction (solid line) (taken from Paik et al., 2015).

The Task and Data

The task that we apply a microgenetic approach to is a large complex office-related task, the Dismal spreadsheet task (Kim & Ritter, 2015). Dismal is a spreadsheet that runs within Emacs and was initially developed to analyze behavioral process models and data (Ritter & Larkin, 1994; Ritter & Wood, 2005).

The Task

The subtasks include: opening a spreadsheet file, saving the file as another name, and completing a complex spreadsheet manipulation by calculating and filling in several blank cells, including five data normalization calculations, five data frequency calculations, ten calculations of length, ten calculations of total typed characters, four summations of each column, and an insertion of two rows to type in the current date and name using Dismal keystroke commands. Together, they can be grouped into the 14 subtasks, as shown in Table 1. More information about the task (e.g., the task environment and the procedure) is available (Kim & Ritter, 2015).

Table 1: The subtasks in the Dismal spreadsheet task.

Subtasks	Keystrokes
(1) Open File	Press `C-x C-f` Type <normalization.dis> ↲
(2) Save As	Press C-x C-w Type JWK.dis ↲
(3) Calculate Frequency (B6 to B10)	Move the point to B6 by using C-p, C-n, C-f, or C-b Press `e` Type `"(/ (* c6 b12) 100.0)"` ↲ Repeat for B7 to B10
(4) Calculate Total Frequency (B13)	Move to the point to B13 Press **e** Type `"(dis-sum b1:b10)"` ↲
(5) Calculate Normalization (C1 to C5)	Move the point to C1 Press `e` Type `"(/ (* 100.0 b1) b12)"` ↲ Repeat for C2 to C5
(6) Calculate Total Normalization (C13)	Move the point to B13 Press **e** Type `"(dis-sum c1:c10)"` ↲
(7) Calculate Length (D1 to D10)	Move to the point D1 Press `e` Type `"(length a1)"` ↲ Repeat for D2 to D10
(8) Calculate Total Length (D13)	Move the point to D13 Press `e` Type `"(dis-sum d1:d10)"` ↲
(9) Calculate Typed Characters (E1 to E10)	Move the point to E1 Press `e` Type `"(* b1 d1)"` ↲ Repeat for E2 to E10
(10) Calculate Total Typed Characters (E13)	Move the point to E13 Press `e` Type `(dis-sum e1:e10)` ↲
(11) Insert Two Rows	Move the point to A0 Press C-u type 2 i r ↲
(12) Type in *Name* (A0)	Press `e` Type in ***`Name`*** ↲
(13) Insert Current Date (A1)	Move the point to A1 Press `e` Type `"(dis-current-date)"` ↲
(14) Save As Printable Format	Press `C-x C-w` Type <normalization-initials.dp >↲

The Data

The data used in this paper is 30 participants' learning performance. A learning session consists of a study session and a test trial. A study session is when a participant used the study booklet to learn. Each study session is limited to 30 minutes of study. A test trial in the learning session is when participants perform the given tasks without the study booklet.

In the first week, participants performed four consecutive learning sessions. On Day 1, participants had a maximum of 30 minutes to study the spreadsheet tasks and then performed the tasks. On Days 2 to 4, participants were allowed to refresh their acquired knowledge from Day 1, using the study booklet, and then performed the tasks.

The task completion time and every keystroke movement were measured by the Recording User Input (RUI) system (Kukreja, Stevenson, & Ritter, 2006). The target participants in this report used a keystroke-based interface to complete the task. The raw data included every keystroke and its time (in ms). This allows us to investigate performance change on a more microgenetic level by examining the time to perform each subtask and unit task during the practice trials.

Microgenetic Analysis of Learning

We next describe the subtasks and then how learning happens by subtask.

Preliminary Analysis of the Subtasks

Table 2 shows the KLM actions in the task based on the instructions. We initially analyzed the recorded performance under the KLM framework as seen in Table 2.

Each subtask has different mental and keystroke operators. The KLM analysis is based on the number of each operator in each subtask according to Eq. 1. It provides us with a basic quantitative baseline prediction of user performance, not performance change. Three practice trials is enough to get to the KLM times. With even more practice performance is faster than the KLM predictions (Card, Moran, & Newell, 1983, p. 285). Approximately half of the tasks are as fast as the KLM on trial 3, and all but one are on trial 4.

Table 2: KLM-GOMS Prediction of Subtasks (in seconds)

Subtasks	Operators				
	M	H	P	K	Time
Sub1	3	1	0	33	19.96
Sub2	3	0	0	26	16.27
Sub3	20	0	0	158	101.26
Sub4	4	0	0	27	18.09
Sub5	20	0	0	169	106.43
Sub6	4	0	0	37	22.79
Sub7	39	0	0	194	143.83
Sub8	4	0	0	27	18.09
Sub9	40	0	0	186	141.42
Sub10	4	0	0	27	18.09
Sub11	2	0	0	39	21.03
Sub12	2	0	0	9	6.93
Sub13	4	0	0	24	16.68
Sub14	3	0	0	25	15.80
Operators	152	1	0	981	
Time	205.20	0.40	0.00	461.07	666.67

Statistical Modeling of Performance Change

The data set used in this paper is longitudinal with repeated measurements for each participant and for each subtask over time. To deal with non-independency in measurements, we choose to use a linear mixed effects model. The response variable in the data set is the task completion time.

In our linear mixed effects model, the fixed effect is the practice trials that is represented as days. As random effects, we had intercepts for participants and subtasks, as well as by-participants and by-subtasks random slopes for the effect of learning trials over time. This statistical model is adequate for the question of interest in this paper, investigating whether different subtasks have differential learning rates by participants over practice trials. Subtasks and participants are completely crossed, and the task time was repeatedly measured from each participant.

Results

Figure 3 shows a preliminary plot of the 14 subtasks. It shows different patterns of performance change across four days of practice trials. The red dashed horizontal lines are the KLM predictions. Figure 3 suggests the practice trials for four consecutive days allow participants to approximately reach an KLM expert performance except for subtasks 7 and 9. Where the KLM predictions seem to be higher than true experts will be, these subtasks have the largest number of mental and keystroke operators (refer to Table 2). This result casts a question as to whether the number of mental operators are over predicted.

We used the *lme4* package (Bates, Mächler, Bolker, & Walker, 2014) in R to conduct a linear mixed effects analysis of the relationship between the response variable and the covariate predictors including fixed and random effects.

We checked the normality assumption of the data. The Q-Q plot of residuals shows that the residuals are not normally distributed. To address this issue, we performed log-transformation of the data. Our linear mixed effects model then meets the assumption of normality of residuals.

To assess the significance of practice trials (day) as a predictor, we looked at the t-value of the fixed effects. The t-value of the slope estimate is large enough. Thus, we can estimate that the predictor is significant since our dataset is fairly large with 1680 observations.

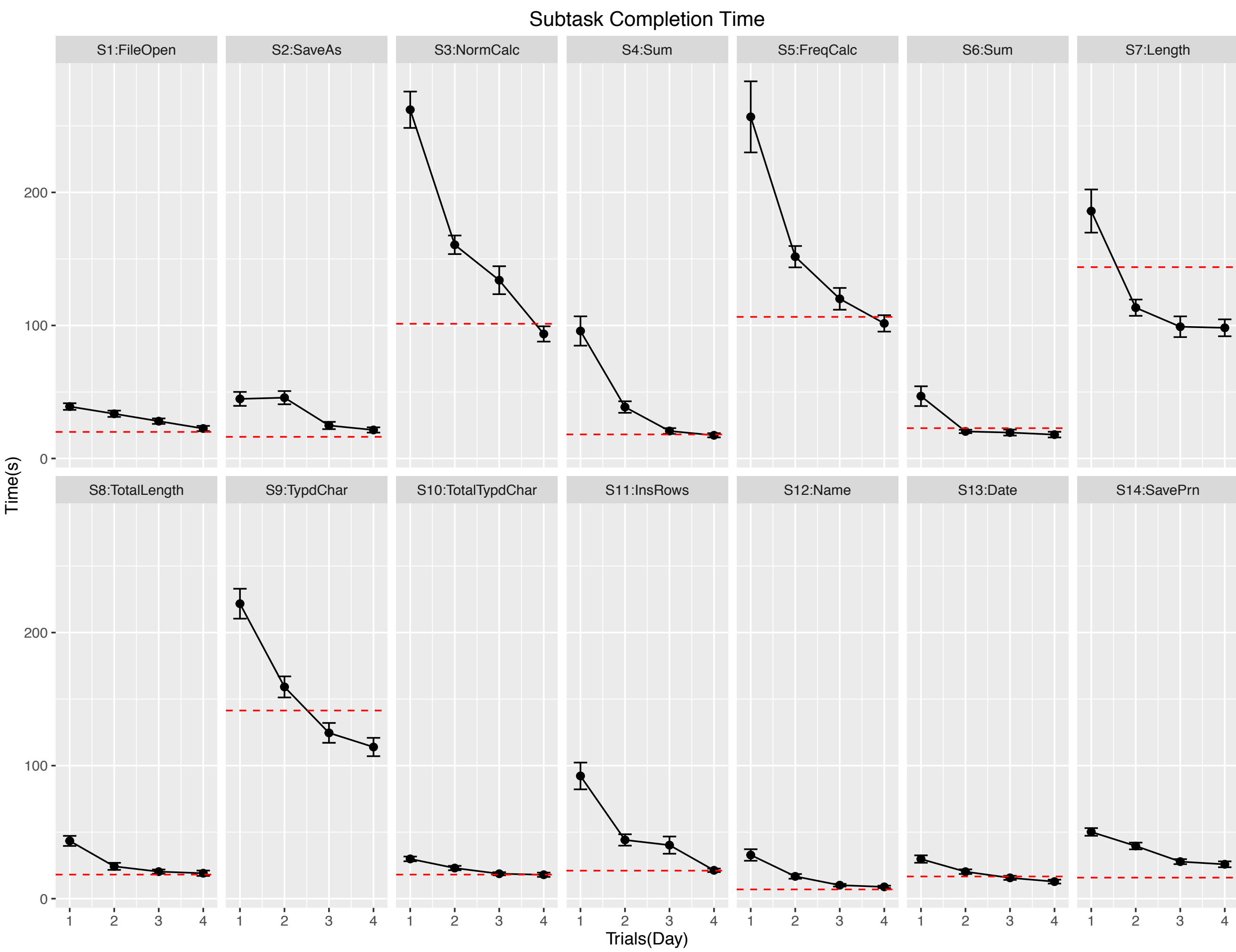

Figure 3: Average subtask completion times (N=30) in seconds with mean (solid black) and SEM (as error bars) for each subtask. The red dashed lines are the KLM predictions for each subtask.

We plotted the data to depict all the task completion times over practice trials by the 14 subtasks (*N*=30), and a linear regression line for each subtask in a log-log coordinates, as shown in Figure 4. There were 48 missing values from 1680 data points (2.9%), but it can be considered that those missing values are acceptable for our model due to the total number of data points.

Besides the fixed effect of practice trials over days, it is of interest to determine how the subtasks differ. We compared two models: one model is a random intercept model both for participants and for subtasks, and the other model is also a random intercept model that has only different intercepts of participants (i.e., without random intercepts of subtasks). The random deviations (residuals, *SD*=0.17) from the predicted values that are not caused by both subtasks and participants increased in the case of the random deviations (*SD* = 0.40) only due to participants. This indicates that the subtasks have an effect on the performance change. By performing ANOVA to compare those two models, we can conclude that there is a statistical significance of the subtask effect, $\chi^2(1) = 2681.4$, $p < 0.001$.

As seen in Figure 4, there exist varying slopes by subtasks, indicating different learning rates by each subtask. With regard to the varying slopes of the subtask effect, we compared the model with random intercepts to an alternative model with random slopes for the subtask. We found there are significant differences in learning rates by the random effect of the 14 subtasks, $\chi^2(2) = 115.59$, $p < 0.001$.

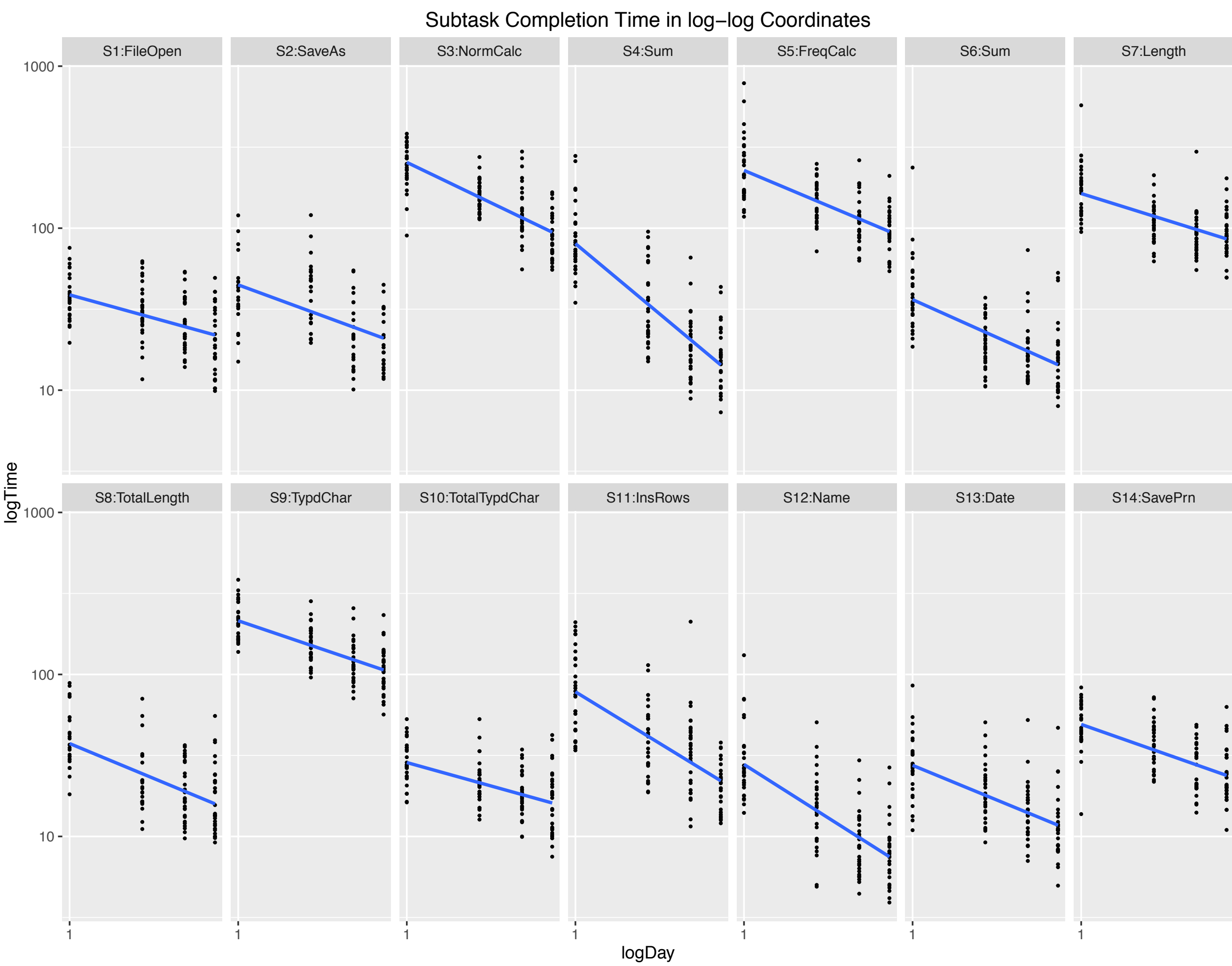

Figure 4. Regression lines with scatter plots for each subtask in a log-log scale.

Discussion and Conclusions

Figure 5 shows differences and similarities in the slope for the predicted time by each subtask. Similar slopes are observed in the subtasks 1, 2, 7, 9, 14, subtasks 3, 5, 6, and subtasks 11, 12. As noted in Table 1, the participants retrieve each keystroke command for the corresponding subtask, such as the unique key commands, C-x C-f, for "Open File", and C-x C-w, for "Save As". In this manner, the operators required for subtask 3 and 5, which are normalization and frequency calculations, are nearly identical. The slopes for the task time predictions are similar as well. However, it is apparent that the slope of subtask 3 is steeper than the slope of subtask 1. It is interesting that the number of operators of either type in a subtask, particularly when there are fewer than 50, is not correlated with learning slopes.

With regard to the subtasks 3, 5 (normalization and frequency calculations), and subtasks 7, 9 (calculating length and typed characters), those subtasks require a large number of keystroke operators in the spreadsheet subtask. However, the number of keystroke operators might not be what influenced learning because there are other subtasks with steeper slopes and fewer keystroke operators. On the other hand, the keystroke skills are learned for four consecutive days of practice. All these subtasks required participants to repeat 10 calculations per practice trial. This can be considered as motor skill practice with a massed training regimen.

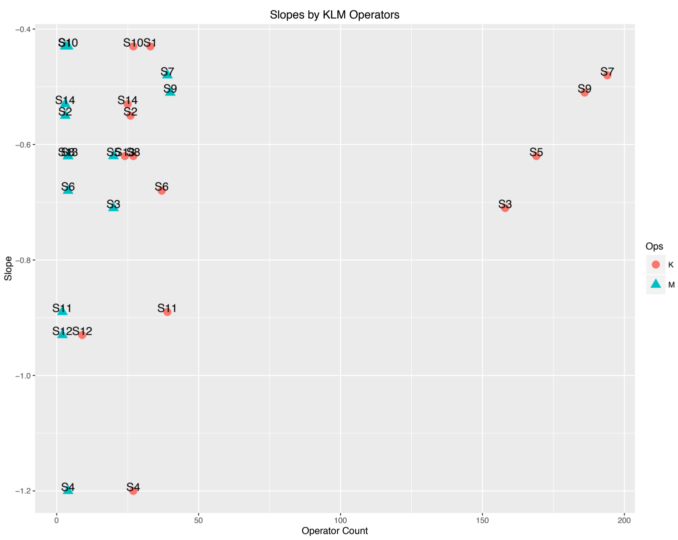

Figure 5: Scatterplot of the varying slopes against operators (Keystroke and Mental). (Lower is greater learning.)

Figure 5 suggests that as mental operators go up, the learning rate goes down, but this seems curious. Regarding mental operators, some subtasks require participants to retrieve a unique keystroke command, and this can lead to higher learning rates. Perhaps these have different effects on learning. For example, to insert two rows, a participant needs to retrieve a declarative memory element, C-u 2 i r (the subtask 11). We can consider that the subtasks 11, 12, and 4 would lead to higher learning rates due to a weak activation of the corresponding element. This notion emphasizes the importance of moving the declarative memory elements to the procedural stage (Fig 1).

This analysis shows that the subtasks vary in learning. We are now analyzing why learning varies so much across subtasks and will be investigating using our existing ACT-R models.

Acknowledgments

This research was supported by grants from the Division of Human Performance Training, & Education at ONR (W911QY-07-01-0004, N00014-10-1-0410, and N00014-15-1-2275). Ysabelle Coutu provided useful inputs.

References

Agre, P., & Shrager, J. (1990). Routine evolution as the microgenetic basis of skill acquisition. In *Proceedings of the 12th Annual Conference of Cognitive Science Society* (pp. 694-701). Cambridge, MA.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*(4), 369-406.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York, NY: Oxford University Press.

Anderson, J. R., Fincham, J. M., & Douglass, S. (1999). Practice and retention: A unifying analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(5), 1120-1136.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum.

Fitts, P. M. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 243-285). New York: Academic Press.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practicie. *Psychonomic Bulletin & Review, 7*(2), 185-207.

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.

Kukreja, U., Stevenson, W. E., & Ritter, F. E. (2006). RUI: Recording user input from interfaces under Window and Mac OS X. *Behavior Research Methods, 38*(4), 656-659.

Moon, J. M., & Fu, W.-T. (2008). A situated cognitive model of the routine evolution of skills. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 935-939). Sage.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, NJ: Lawrence Erlbaum.

Paik, J., Kim, J. W., Ritter, F., & Reitter, D. (2015). Predicting user performance and learning in human-computer interaction with the Herbal compiler. *ACM Transactions on Computer-Human Interaction, 22*(5), 1-26.

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

Ritter, F. E., Baxter, G. D., Kim, J. W., & Srinivasmurthy, S. (2013). Learning and retention. In J. D. Lee & A. Kirlik (Eds.), *The Oxford handbook of cognitive engineering* (pp. 125-142). New York, NY: Oxford University Press.

Ritter, F. E., & Larkin, J. H. (1994). Developing process models as summaries of HCI action sequences. *Human-Computer Interaction, 9*, 345-383.

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.

Siegler, R. S. (2006). Microgenetic analyses of learning. In *Handbook of child psychology* (pp. 464-510): Wiley.