Proceedings of 25th Army Science Conference, V43, Orlando, FL. Nov. 27-30, 2006. CD Rom
(page number is not available at this time).

# DETERMINING LEARNING LEVEL AND EFFECTIVE TRAINING TIMES USING THERMOGRAPHY

J. Kang, J. A. McGinley, and G. McFadyen
Center for Advanced Vehicular Systems, Mississippi State University
Mississippi State, MS 39762-9542

K. Babski-Reeves
Department of Industrial and Systems Engineering, Mississippi State University
Mississippi State, MS 39762-9542

## ABSTRACT

The goal of this study was to demonstrate a measurable correlation between facial temperature changes and the human learning process. This study also introduces a method for determining sufficient training time by monitoring the cognitive load of the learner as indicated by changes in facial temperature. A non-invasive method using an infrared camera was employed to detect nose temperature changes during performance of a novel alpha-numeric task. Nine participants (5 males and 4 females) completed 7 blocks of the novel task. Changes in nose temperature, reaction time, accuracy on task, and subjective perceptions of mental workload were collected. In general, reaction time and subjective perceptions of workload decreased across blocks, while nose temperature and accuracy were found to increase. The majority of changes were found to occur from the first block to the second, indicating that a significant amount of learning had taken place. Significant performance improvements were not found beyond the first block, though subjective perceptions of workload consistently decreased across blocks. Strong correlations were found between each of the dependent variables indicating that thermal readings, MCH ratings, and performance are drawing on the same constructs (mental workload and learning). Given these findings, thermography may be useful in defining sufficient training times for learning novel tasks, thereby improving training quality and reducing training costs.

## 1. INTRODUCTION

The U.S. Army trains about 80,000 new recruits annually, the vast majority of whom are recent high school graduates. The U.S Army has a responsibility to train and educate soldiers to execute their mission, while maintaining soldier safety, though most new recruits do not have experiences with military weapons and operations.

Training is a critical component of military operations because there is no military strength without well-trained soldiers. Changes in military operations and engagement strategies have necessitated the need for revised training approaches, such as virtual training from remote locations.

The purpose of training is to facilitate learning skills and knowledge required to perform specific tasks (Swezey and Pearlsten, 2001). The New Oxford American Dictionary (2001) defines learning as “the acquisition of knowledge or skills through experience, practice, or study, or by being taught.” Training effectiveness is often assessed using post training questionnaires to identify whether skills or knowledge have been assimilated. While traditional evaluation methods may be effective in evaluating knowledge-based learning, they are not as useful for evaluating task-based skill learning. Further, appropriate training times, for traditional and new training strategies, are not well documented in the literature.

The learning process imposes mental workload on the learner during skill acquisition, though the magnitude of mental workload levels is dependent upon individual differences (Jex, 1988). Previous research indicates a relationship between mental workload (and factors related to mental workload) and learning. For example, Wickens and Hollands (2000) found that reducing mental workload results in improved learning.

Performing novel tasks imposes increased workload levels and requires more reasoning, rehearsal, and mental effort (Baddelely, 1986). Training and practice contribute as psychophysiological influences in reducing a person's mental workload (Jex, 1988).

Techniques for monitoring learning progression during training may be useful in defining training times and assessing the quality or extent of learning. Previous researchers have suggested that there is a point when a necessary and sufficient amount of training time has elapsed for a novice to acquire a skill (Kaber *et al*, 2001; Young and Stanton, 2001), and this level is related to autonomic nervous system (ANS) responses, primarily body temperature. Various measurable biological quantities related to the ANS (such a blood flow changes) are used to estimate arousal levels (Boucsein, 1993), which is closely related to mental workload and learning.

Thermography is a non-invasive technique used to measure temperature at the surface level. Learning requires a certain level of cognitive arousal, which is also closely related to other cognitive concepts such as anxiety, attention, agitation, stress, and motivation. Sensations such as stress, anxiety or fatigue can bring about measurable levels of change in a person's body temperature (Genno *et al*., 1997), detectable by thermal imaging. Veltman and Vos (2005) studied the relationship between facial temperature and workload using thermography. However, very little research exists to describe the relationship between mental workload and thermography, and between training time and learning. Therefore, the objective of this study was to quantify the relationship between thermography, mental workload, and training time (or learning).

## 2. METHODOLOGY

### 2.1 Experiment Design

A one-factor, within-subject design was used to assess learning (experimental block) on thermal readings of the nose, reaction time, response accuracy, and subjective workload assessment ratings.

### 2.2 Independent Variable

The independent variable for this study was experimental block, associated with an alphabet arithmetic task (Logan and Klapp, 1991) used to assess learning a novel task. This task requires participants to verify equations of the form C+2=?, D+3=?, and E+4=?. Participants count through the alphabet one letter at a time until the answer to the equation is determined. Only the letters C, D, or E and the numbers 1—4 are used in this task. The experiment consisted of 7 task blocks, each 6 minutes and 24 seconds in length, with a 3 minute rest period between blocks. Each block included 8 question sets consisting of 12 questions (one block = 96 questions). The same 12 questions are used in each set, though question order was randomized across sets. A single question (D+4=?) appeared in the same place of each set to allow comparison of thermal readings at a consistent point across each set. Each participant had a 4-second time limit to respond to each question. Question presentation time was synchronized with the thermal camera to aid in data analysis.

### 2.3 Dependent Variables

A MikronScan 7200V Thermal Camera (Mikron Infrared, Inc., Oakland, NJ) captured facial thermal images at a rate of 1Hz for the entire testing session. The camera was positioned to the left of the computer screen approximately three feet from the participants face. The camera was focused to fully visualize the face and neck of each participant.

Individual image frames were analyzed by placing a region of interest (ROI) over the nose (Figure 1) using MikroSpec R/T software. Standardization of ROIs within participant frames was ensured by using participant landmarks and standardization of ROIs across participants was ensured by capturing images from the midpoint of the nostrils for the nose. Though thermal data was available for the entire test session, only the images associated with a single question across each block were analyzed (details provided under Independent Variable). The maximum values for each frame were recorded, and the averages of these temperatures for the period of time associated with the single question were used in data analysis.

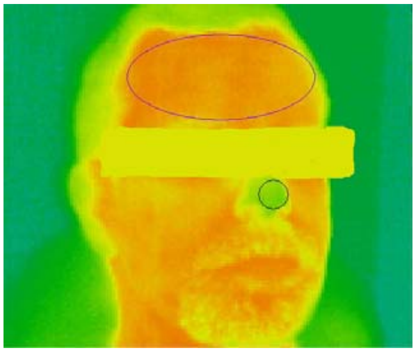

Figure. 1. Example ROI for a single thermal image. In this study only the circular nose ROI was analyzed.

Subjective ratings of mental workload were captured using the Modified Cooper Harper Scale (MCH) (Wierwille and Casali, 1983). The MCH is a uni-dimensional scale in which a series of questions directly leads to a single rating of mental workload that ranges form 1 (very easy) to 10 (impossible). The advantages of the MCH include ease of use and reduction in administrative work. This scale was used because it was specifically designed to be used in tasks that are perceptual and communicative in nature (Wierwille and Casali, 1983). Participants completed the MCH after each experimental block.

Participant reaction time (RT) and response accuracy were also recorded. RT was defined as the time of question onset to the participant's response. Accuracy was defined as the percentage of correct questions answered. Participants failing to answer the question within the allotted time frame received an incorrect response score. Accuracy for each trial was computed and compared.

### 2.4 Participants

Nine participants (5 males and 4 females) ranging from 18 to 26 years completed the experimental protocols. All participants are college students with similar education levels.

### 2.5 Procedure

Participants were initially provided with a verbal and written description of the study, its objectives and procedures, and completed informed consent documents approved by the Mississippi State University Institutional Review Board for Research Involving Human Subjects. Participants completed a demographic questionnaire, and were seated at the experimental station. Thermal readings began just prior to the first question, and continued to run throughout. Participants completed each of the 7 task blocks, followed by a 3 minute rest period during which they completed the MCH. At the end of the final experimental block and MCH rating, participants were monetarily compensated for their participation.

### 2.6 Data Analysis

Appropriate descriptive statistics were computed for each dependent variable (e.g., means, standard deviations, frequency counts). A repeated measures ANOVA was used to assess the effects of experimental block on facial thermal readings, participant reaction time (RT), response accuracy, and MCH ratings of mental workload. Gender differences were not considered. Tukey's HSD post hoc tests were used where appropriate. Correlations between each of the dependent variables were also calculated. All findings were considered significant at an alpha of 0.05. All analyses were performed on SAS 9.1 statistical software.

## 3. RESULTS

Descriptive statistics for each of the dependent variables are provided in Table 1. In general, subjective workload ratings and RT decreased across experimental blocks, while response accuracy and nose temperature increased over blocks.

All the dependent variables were significantly affected by block. Nose temperature readings were found to be significantly lower in block 1 than the other blocks, though no other differences were found (Figure 2). This same finding held true for response accuracy. All blocks were found to differ from each other, except for blocks 5, 6, and 7, for

MCH ratings, meaning that subjective ratings of mental workload decreased significantly until the 5th block where no other statistical differences in workload ratings were found. Similar trends were found for RT. RT was found to decrease significantly across the first three blocks, though no other significant differences between blocks were found.

Strong correlations were found between each of the dependent variables (Table 2). Correlation coefficients were found to be greater than 0.88 (in either direction), with p-values significantly less than 0.05. The largest significant correlation was found between reaction time and accuracy at 0.98. Significant correlations with nose temperature ranged from -0.88 to 0.96, indicating that thermal imaging was sensitive to learning and workload levels.

Table 1. Descriptive Statistics for the Dependent Variables (values are in mean (std dev))

| | Dependent Variable | | | |
|---|---|---|---|---|
| Block | Nose Temperature (°C) | Reaction time (sec) | Response Accuracy (%) | MCH Ratings |
| 1 | 34.12 (2.31) | 1.9701 (0.22954) | 91.67 (0.06) | 4.11 (0.93) |
| 2 | 34.55 (2.29) | 1.6971 (0.28460) | 95.78 (0.04) | 3.44 (0.53) |
| 3 | 34.78 (1.94) | 1.5888 (0.23668) | 97.33 (0.03) | 3.11 (0.60) |
| 4 | 34.77 (2.32) | 1.5258 (0.22922) | 97.00 (0.03) | 2.56 (0.73) |
| 5 | 34.79 (2.30) | 1.5130 (0.23028) | 98.00 (0.01) | 2.11 (0.60) |
| 6 | 34.77 (1.98) | 1.4497 (0.25165) | 98.67 (0.01) | 1.89 (0.78) |
| 7 | 34.99 (2.13) | 1.4776 (0.32103) | 98.00 (0.02) | 1.44 (0.53) |

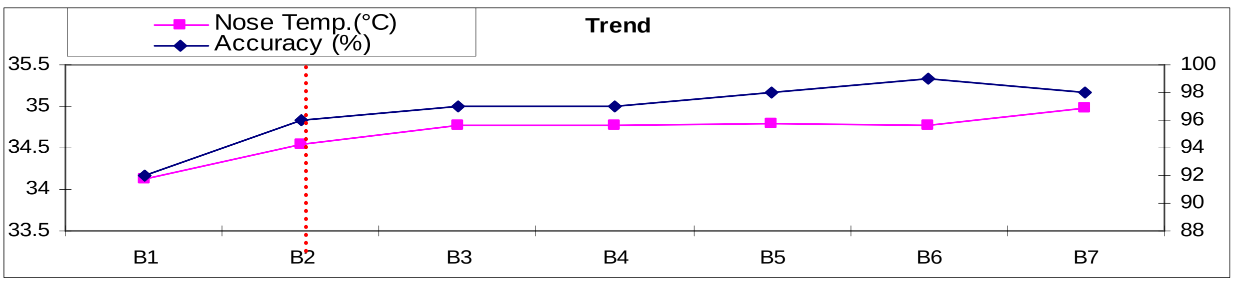


Figure. 2. Nose temperature and accuracy trends across treatment blocks.

Table 2. Correlation Coefficients

| Variable | NT | Acc | RT | MCH |
|---|---|---|---|---|
| NT | 1 | | | |
| Acc | 0.96 | 1 | | |
| RT | -0.95 | -0.98 | 1 | |
| MCH | -0.88 | -0.88 | 0.91 | 1 |

## 4. DISCUSSION

The goal of this study was to assess the efficacy of thermography for identifying minimum training times for the learning of novel tasks. Thermogaphy correlated strongly with traditional (objective) evaluation methods (accuracy and reaction time), supporting the use of thermography as an objective tool to assess learning and workload.

Nose temperatures were initially affected by exposure to a novel task, potentially due to increased workload during learning. However, changes in thermal readings were not sustained, suggesting that most of the learning/training process had taken place. This hypothesis was supported by the ceiling effect found for accuracy and reaction time, though these leveled off at a later time interval. Nose temperature, reaction time, and response accuracy reached at point where significant improvements were no longer observed at blocks 1, 3, and 5 respectively. Therefore, while learning effects were significant early, training briefly beyond this point can improve performance. Extended training however, would not result in performance improvements.

Nose temperature was found to increase or remain constant across the test session. Though the ANS response can be considered automatic, the physiological responses associated with changes in the ANS signal may not be immediately observable. Therefore, it was expected that despite a reduction in workload, nose temperature readings may remain elevated. The current study was not long enough in length to identify the long term trend in nose temperature, though it is assumed that over time, as workload continued to decrease or level off, nose temperature readings would also. Further studies are needed to support this hypothesis.

The leveling off of nose temperature readings raises an issue on how to interpret thermal image findings. We propose that if significant differences are found between time intervals, then learning and workload are still occurring, thus additional training is needed for skill mastery. The point at which no differences are detected in thermal images is the point at which sufficient training time has been provided to learn a new skill. Continued practice would result in improved performance (as was found here), though the improvements in performance are to a lesser degree than when learning is occurring (i.e., improvement is not statistically significant). Support for this hypothesis can be found in the workload, performance, and demand relationship described by de Waard (1996). This relationship states that under low and high task demands, workload will be high and performance low. Between the two demand extremes workload levels will decrease, while performance levels increase (Figure 3). When performing a novel task, task demands are high and should result in poor performance and higher workload levels, as was found in this study. As persons become more proficient at the task, then task demand levels should decrease, and performance and workload levels should follow the curve from right to left.

Though MCH ratings were found to differ significantly across most blocks, only block 1 is categorized as “difficult” according to the scale authors (Wierwille and Casali, 1983). Ratings of “3” and less are considered easy and are thought to impose minimal workload on the operator. This finding again supports that sufficient training had occurred following the first experimental block for mental workload levels to adjust to a level at which performance begins to be maximized.

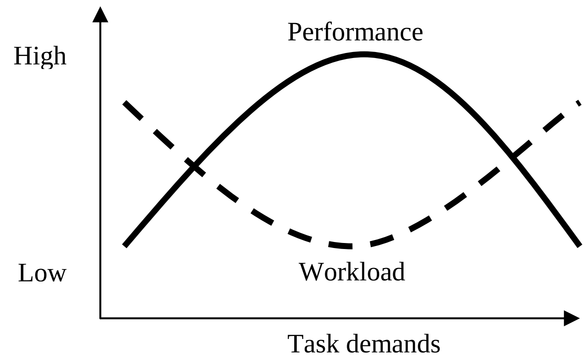


Figure 3. Workload, performance, and task demands relationship (adapted from de Waard, 1996)

The utility of using thermography to identify sufficient training times was illustrated by the results of this study. As it is not always convenient or feasible to obtain participant perceptions through subjective ratings, thermography may provide a non-invasive methodology to assess training times. Further research is needed to fully understand the use of the thermography in assessing learning and defining training times.

## REFERENCES

Baddeley, A.D. (1986). *Working memory*. Oxford, UK: Oxford University Press.

Boucsein, W. (1993). Psychophysiology in the work place-Goals and methods. In P. Ullsberger (Ed.) *Psychophysiology of mental workload* (pp. 35-41). Berlin: Bundesanstalt fur Arbeitmedizin.

De Waard, D. (1996). *The measurement of driver’s mental workload*. PhD Dissertation, University of Groningen, Netherlands.

Genno, H., Ishikawa, K., Kanbara, O., Kikumoto, M., Fujiwara, Y., Suzuki, R., *et al.* (1997). Using facial skin temperature to objectively evaluate sensations. *International Journal of Industrial Ergonomics, 19,* 161-171.

Jewell, E.J., & Abate, F. (Eds.). (2001). *The new Oxford American dictionary (1st ed.).* New York: Oxford University Press.

Jex, H. R. (1988). Measuring mental workload: Problems, Progress, and Promises. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 5-39). New York: North-Holland.

Kaber, D. B., Riley, J. M., Tan K., & Endsley, M. R. (2001). On the design of adaptive automation for complex systems. *International Journal of Cognitive Ergonomics, 5*(1), 37-57.

Logan, G.D., & Klapp, S.T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity?. *Journal of Experimental Psychology: Learning Memory, and Cognition. 17*(2), 179-195.

Veltman, H.J.A., & Vos, W.K. (2005). Facial temperature as a measure of operator state. *11$^{th}$ International conference on Human-computer interaction*, Las Vegas.

Wickens, C.D., & Hollands, J.G. (2000). *Engineering psychology and human performance (3rd ed.).* Upper Saddle River, NJ : Prentice Hall.

Wierwille, W.W., & Casali, J.G. (1983). A validated rating scale for global mental workload measurement applications, *Proceedings of the Human Factors Society 27$^{th}$ Annual Meeting*, Santa Monica, CA, 129-133.

Swezey, R.W. & Pearlsten, R.B. (2001). Selection, training, and development of personnel, In G. Salvendy (Ed.), *Handbook of Industrial Engineering: Technology and Operations Management ( 3$^{rd}$ ed.)* (pp. 920-947). New York, NY: John Wiley& Sons, Inc..

Young, M.S. & Stanton, N.A. (2001). Mental workload: Theory, measurement, and application. In W. Karwowski (Ed.), *International Encyclopedia of Ergonomics and Human Factors* (pp. 507-509). New York, NY: Taylor and Francis.