Identifying Automation Opportunities in Life Science Processes through Operator Task Modeling and Workload Assessment

Manida Swangnetr[ab] , David Kaber[c], Ellen Vorberg[a], Heidi Fleischer[a] and Kerstin Thurow[a]

[a]Center for Life Science Automation,
University of Rostock, Rostock, 18119 Germany
[b]Back, Neck and Other Joint Pain Research Group, Department of Production Technology,
Khon Kaen University, Khon Kaen, 40002 Thailand

[c]Edwards P. Fitts Department of Industrial and Systems Engineering,
North Carolina State University, Raleigh, NC, 27695-7906, USA

ABSTRACT

In an effort to automate manual life science processes for high throughput and accuracy, we previously observed that perceived operator workload could be used to identify taxing tasks as targets for robotics. However, we also observed that other factors, including task time and step count, might influence workload. The objective of the present research was to determine whether technician perceptions of workload were driven by process method characteristics, specifically duration, number of steps, and numbers of motor and cognitive operations. Confirmation of influence of these characteristics on perceived workload was expected to provide further direction for automation development for specific methods. A hierarchical task analysis was prepared for a mercury analysis process and revealed various methods for accomplishing goals. Methods included sequences of operations, which were subsequently classified as perceptual, motor or cognitive in nature by using GOMS methodology (Goals, Operators, Methods, and Selection rules). A field study was conducted with three lab technicians completing the mercury analysis process in three replications. Perceived workload for each method was collected using the NASA-Task Load index (TLX). Significant positive correlations were found between method times and operation counts determined based on GOMS models with technician overall TLX ratings. Motor, cognitive and combinations of both operator counts were also correlated with TLX physical, mental demand and effort ratings, accordingly. In general, longer duration methods, including weighing, tuning and pipetting steps, appear to pose high workload for technicians and represent priority targets for automation. Furthermore, a sequence of recollection and planning operations as part of a pipetting task posed the greatest sustained cognitive load for technicians and may represent an opportunity for use of advanced robotic technology with capacity to act as an assistant to technicians.

Keywords: Cognitive Workload, Hierarchical Task Analysis, GOMS, Human-Automation Interaction, Life Science Processes

INTRODUCTION

One research objective of life science laboratories is to develop novel analytical methods to screen construction material samples for toxic chemicals that are hazardous to the environment in the event of unmanaged disposal. In development of such processes, laboratory technicians are initially required to perform manual protocols on samples of known chemical concentrations in order to achieve an optimal technique. The Center for Life Science Automation (CELISCA) at University of Rostock (DE) aims to develop automated systems to perform methods as part of such manual protocols in order to meet high throughput demands and test accuracy requirements as well as ensure technician safety and reduce cognitive workload. In a previous study (Swangnetr et al., current volume), we investigated technician perceived workload in a mercury content determination process for waste wood materials (see Fleischer and Thurow (2013) for process description). This investigation was intended to provide a basis for automating manual procedures as part of the mercury analysis. Findings indicated assessment of technician perceived load could be used to effectively identify taxing tasks for targeting applications of automation. Prototyping of an automated workstation as part of the process and subsequent analysis of technician workload revealed significantly lower cognitive load in manual tasks performed with the automation, as compared with purely manual process activities. However, lab technician workload ratings for task types, with similar sequences of operations and information requirements, were found to be inconsistent between manual and automated systems. This result may be due to other factors, including task completion time and number of steps, influencing workload. These factors may represent additional bases for targeting process automation.

On this basis, the present study sought to model life science lab technician manual performance of the mercury content analysis process and to assess technician cognitive workload as bases for effectively directing process automation efforts. The task modeling was integrated with subjective ratings of workload to determine whether technician perceptions of load were driven by process method characteristics, specifically duration, number of steps, number of perceptual and motor operations, and number of cognitive operations. Confirmation of influence of these characteristics on technician workload was expected to motivate use of both workload and task information for developing automation for specific process methods. The overarching goal of this work was to promote greater process efficiency and accuracy in analysis of aged, treated-wood construction materials as a basis for safe disposal.

CONSTRUCTION OF COGNITVE TASK MODEL

Goals, Operators, Methods, and Selection rules (GOMS) is a cognitive task analysis methodology, originally developed by Card, Moran and Newell (CMN; 1983). A GOMS model consists of descriptions of “methods” needed to accomplish specified “goals”. The methods are series of steps, or “operators” (hereafter referred to as “operations”), performed by the user. A method can also include sub-goals to be accomplished with return to the primary method-for-goal. Therefore, methods have a hierarchical structure in GOMS analysis. “Selection rules” are used to choose the appropriate method, if there is more than one method to accomplish a goal. Users must decide on a rule, based on task circumstances, in order to effectively accomplish the goal. GOMS was developed for decomposition of procedural tasks as performed by expert operators. GOMS has been widely used in human-computer interaction domain (e.g., Polson and Kieras, 1985; Gray et al., 1992; Salvucci and Lee, 2003). The methodology has also been successfully applied to other domains, for example aviation (Foyle et al., 2005) and human-robot interaction (Drury et al., 2007). Since the development of CMN-GOMS, computer-based forms of the methodology have been created, including the GOMS Language (GOMSL; Kieras, 2005). We have previously used GOMSL for modeling human teleoperation of a rover in a simple path-following task (Kaber et al., 2011); however, we have not used this language to represent complex manual and automated procedures in life science processes. Other task analysis methods, such as hierarchical task analysis (HTA; Diaper, 1989) or goal-directed task analysis (Endsley, 1993), do not provide detailed structure comparable to GOMSL for task representation, identification of specific types of perceptual, motor and cognitive operations as part of task performance, or the capability to identify the flow of operator information processing.

A HTA was initially prepared on the manual mercury analysis process to identify user goals and sub-goals. Methods for accomplishing goals were also identified. The HTA was based on review of a standard operating procedure (SOP), retrospective think-aloud protocols with life science lab technicians (while viewing videos of the manual process), and interviews with process experts on task objectives and potential errors. Using the GOMS methodology, each task method was further detailed in terms of sequences of operations. (For operations that could not be coded using the conventional set of GOMS operators, user-defined operators were used.) Operations were subsequently classified as perceptual (P), motor (M) or cognitive (C) in nature. Table 1 shows the goals, sub-goals and methods, obtained from the HTA for the mercury analysis process. The table also includes numbers of perceptual, motor and cognitive operators for each method, obtained using the GOMS methodology. (Note that the number of operators presented here was based on actual process observations.)

Table 1: Mercury analysis process goals and methods, along with associated numbers of perceptual, motor and cognitive operators

Goal: Sub-goal	**Method**		**#Perceptual operators**	**#Motor operators**	**#Cognitive operators**
Sample preparation:					
Weigh sample	1.1	Prepare weighing workstation	7	13	17
	1.2	Perform weighing of sample	204	150	583
	1.3	Clean-up weighing workstation	5	11	10
Pre-digest sample	2.1	Prepare pre-digestion workstation	6	23	12
	2.2	Perform pre-digestion (pipetting) of sample	96	66	409
	2.3	Clean-up pre-digestion workstation	3	6	7
Digest sample	3.1	Perform microwave digestion	28	50	97
Dilute sample solution	4.1	Prepare sample solution dilution workstation	6	18	19
	4.2	Prepare microwave digested sample	3	20	8
	4.3	Perform diluting (pipetting) sample solution	141	86	513
Calibration solution preparation	5.1	Prepare calibration solution workstation	13	20	31
	5.2	Prepare (pipetting) high standard	54	49	244
	5.3	Perform pipetting of calibration solution	61	49	246
Inductively coupled plasma-mass spectrometry (ICP-MS) analysis	6.1	Prepare ICP-MS machine	15	20	21
	6.2	Start measurement control program	16	19	67
	6.3	Check performance report	18	36	109
	6.4	Tune device and prepare batch	63	103	299
	6.5	Arrange solutions in auto sampler	5	33	52
	6.6	Perform mercury analysis	6	8	27
	6.7	Analyze results	82	123	385
Transportation	7.1	Transport sample to lab	3	10	6
	7.2	Transport solutions to ICP-MS machine	3	8	6

With respect to the GOMS model, technicians typically performed certain tasks as part of the analysis process by following only a subset of the methods described in the model. For example, technicians performed pipetting by either using long-term memory (LTM) or looking at SOPs for specific dissolution procedures. For tasks involving a variation of looping on a set of GOMS operators (i.e., different a number of process adjustments in each trial), the average number of adjustments across trials was used to quantify the model. In this study, there were, on average, five process adjustments when performing sample weighing and/or device tuning methods. Table 2 presents an example GOMS operation sequence (model code) for a portion of the mercury analysis process in which technicians performed pipetting of a “standard” solution for system calibration. The Table shows the specific methods to the goal, the specific cognitive, perceptual and motor operations, as well as their counts, when occurring in sequence.

Table 2: Example of GOMS operation sequence for mercury analysis process

Goal: **Method**	**Operation**	**Description**	**Class**	**#**
Perform pipetting of mercury standard:				
Perform pipette using LTM	Recall_LTM_item	Recall and store in WM <current pipetting task>	C	2
	Decide:	LTM for pipetting parameters is Complete? [Yes]	C	1
	Store	Store <solution name>, <pipetting volume> and <pipetting destination> in WM	C	3
	Delete	Delete <current pipetting task> from WM	C	1
Plan pipetting solution	Think _of	Calculate plan of pipetting solution volume	C	1
	Store	Store <solution volume plan> in WM	C	1
	Store	Store <current solution volume> in WM	C	1
	Decide:	Current solution volume in None? [No]	C	1
Get pipette	Look_at	Current pipette	P	1
	Store	Store <pipette label> in WM	C	1
	Decide:	Have current pipette? [Yes]	C	1
	Decide:	If <pipette label> is correct for <current solution volume>? [No]	C	1
	Do_UserDefined	Put current pipette back	M	1
	Look_for_object and_store	Right pipette for current solution volume and store in WM < pipette label>	P C	1 1
	Do_UserDefined	Get right pipette for current solution volume	M	1
	Delete	Delete <pipetting label> from WM	C	1
Adjust pipette volume	Look_at	Pipette volume interface	P	1
	Store	Store <pipette volume> in WM	C	1
	Decide:	If <pipette volume> is_equal_to <current solution volume>? [No]	C	1
	Do_UserDefined	Adjust pipette volume	M	1
	Delete	Delete <pipetting volume> from WM	C	1
Perform pipetting task	Look_at	Current pipette tip	P	1
	Store	Store <pipette tip status> in WM	C	1
	Decide:	Have pipette tip? [No]	C	1
	Look_at	Pipette tip	P	1
	Do_UserDefined	Attach pipette tip	M	1
	Delete	Delete <pipette tip status> from WM	C	1
	Look_at	Hg standard solution	P	1
	Do_UserDefined	Asparate Hg standard solution into pipette	M	1
	Look_at	Pipette destination	P	1
	Do_UserDefined	Dispense liquid to Pipette destination	M	1
	Think_of	Can use the same tip for next solution?	C	1
	Decide:	Can use the same tip for next solution? [No]	C	1
	Look_at	Pipette tip trash	P	1
	Do_UserDefined	Discard tip	M	1

STUDY PROCEDURE

A field study was conducted with three professional lab technicians. Each technician was asked to complete the mercury analysis process with three replications. Prior to performing the process, technicians completed pairwise rankings of all six workload demand components of the NASA-Task Load index (TLX; Hart and Staveland, 1988). The technicians selected those TLX components that they considered to be the most important contributors to workload in the mercury analysis process. They were then informed how the process was broken-down into methods and steps. An analyst used a custom Android-based application with a smartphone (based on Zhang et al. (2013) system platform) to record the steps performed by technicians, times, and technician ratings of workload demands for each method.

HYPOTHESES

It was hypothesized that method times, collected using the smartphone, would be correlated with lab technician overall TLX ratings (Hypothesis (H)1). Subjective ratings of workload were expected to increase as task time increased. Prior studies have examined changes in workload over the duration of task performance (Haga et al., 2002). In general, task duration has been found to affect workload responses. Mental fatigue appears to accumulate rapidly under higher demand conditions. Our previous study (Swangnetr et al., 2012) also revealed a significant positive linear association of life science technician overall workload scores with sub-process step time. Longer duration tasks were subsequently considered to be priority targets for automation and potential reduction of technician workload.

Operation counts obtained from GOMS models were also expected to be positively correlated with perceived overall workload ratings (H2). Although Kieras (2006) stated that the relationship between operation counts and user workload is not certain, GOMS model method execution time is determined based on the number of operations that must be executed to accomplish a method. Therefore, if a correlation between method time and perceived workload is observed (i.e., in addressing H1), a greater number of operations would be expected to produce higher workload ratings. Certain TLX demand component ratings were also expected to be associated with specific classes of GOMS operations (H3). Motor and cognitive operation counts were expected to be correlated with TLX physical and mental demand components, respectively. A combination of motor and cognitive operator counts was also expected to be correlated with overall effort ratings. This hypothesis was based on the definitions of the TLX subscales (Hart and Staveland, 1988) identifying mental and physical demand as being rated based on levels of mental and physical activity required by a task, respectively. According to the TLX, perceived effort is based on the level of task difficulty in terms of both mental and physical aspects. Support for H2 and H3 would confirm the use of GOMS models for representing life science technician workload. Consequently, models could be used to further facilitate identification of process automation targets beyond high workload methods and at the level of specific sequences of operations. (TLX ratings (alone) allow for taxing method identification.)

ANALYSIS AND RESULTS

Correlation analyses were conducted to determine the degree of association of method characteristics with perceived workload. Results (see Table 3) revealed significant positive correlations between method times (recorded with the smartphone) and operation counts obtained from GOMS models with technician overall TLX ratings. Motor and cognitive operator counts for process methods were also significantly correlated with TLX physical and mental demand component ratings, respectively. Technician perceptions of effort were also found to be positively correlated with the sum of motor and cognitive operation counts. Roughly between 32 and 46% of the variability in technician perceived workload was explained by the GOMS model characteristics.

Table 3: Correlations between method characteristics and NASA-TLX workload ratings (* - significant at alpha = 0.05 level)

Method characteristics	**TLX rating components**	**Correlation**	
Method times	Overall workload demand	r = 0.439	p <.0001*
GOMS operation counts	Overall workload demand	r = 0.385	p <.0001*
Motor operation counts	Physical demand	r = 0.319	p <.0001*
Cognitive operation counts	Mental demand	r = 0.373	p <.0001*
Motor and cognitive operation counts	Effort	r = 0.462	p <.0001*

AUTOMATION TARGETS IDENTIFICATION

In line with expectation (H1), method times were found to be correlated with lab technician workload ratings. The longer duration methods appeared to pose high workload for technicians. An ANOVA model was subsequently structured to identify those methods with significantly longer durations than others. Results (see Figure 1) revealed significant differences in completion time among methods ($F = 82.4762$, $p<0.0001$). Post-hoc results indicated a set of methods, including weighing samples (ID#1.2), device tuning and batch preparation (ID#6.4), and pipetting samples and solutions (ID#2.2, 4.3, 5.2 and 5.3), to require significantly longer completion times as compared with all other methods (again see Figure 1). On this basis, these methods were considered to be targets for automation in order to reduce technician workload.

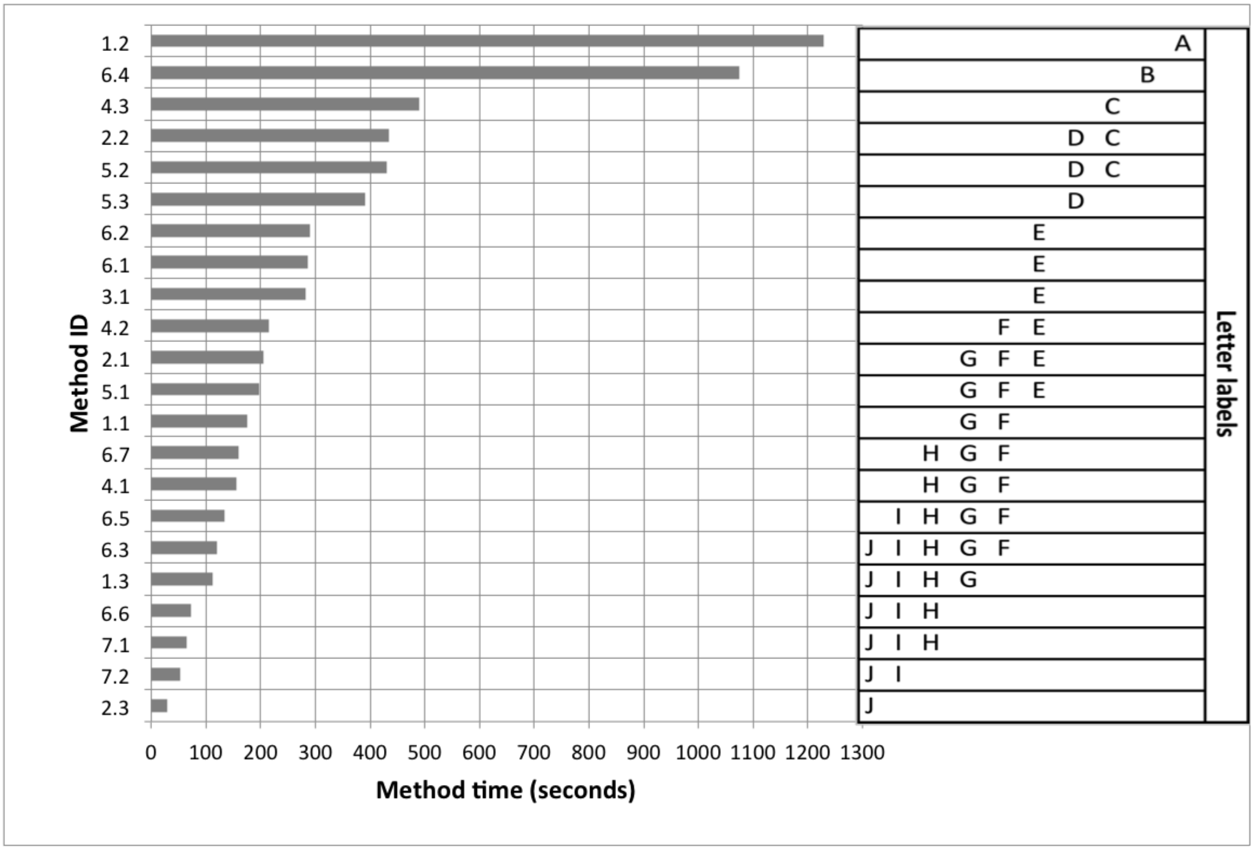

Figure 1. Post-hoc test results on completion times for each method (means with different letter labels are significantly different with $p<0.05$).

In line with H2, operator counts obtained from the GOMS model were found to be correlated with lab technician overall TLX ratings. Motor operations, cognitive operations and the sum of counts for both types of operations were also correlated with TLX physical demand, mental demand, and effort component ratings, accordingly. These results were in support of H3. The findings also indicated GOMS models could be useful for identifying types of operations as well as sequences represent opportunities application of automation and/or advanced robotic technology in order to reduce technician workload. Among the TLX workload demand components, technicians rankings revealed performance and mental demand to be perceived as the first and second greatest contributors to workload in the mercury analysis process, respectively. When examining the GOMS model, it was also found that all methods required frequent use of cognitive operations for accomplishing goals. On these bases, cognitive operations can be considered an important aspect of workload in the mercury analysis process. Consequently, we identified long duration task methods that posed the greatest cognitive requirements for technicians. Weighing samples, device tuning and pipetting methods were inspected for sequences of cognitive operations and counts in order to identify the greatest sustained cognitive load on operators within the methods.

Table 4 shows the maximum number of consecutive cognitive operations occurring in each method. It can be seen that a sequence of recall and planning operations as part of pipetting task performance represented the greatest sustained cognitive load on technicians. This sequence required technicians to perform complex mental processing, including Think_of “calculate plan of pipetting solution volume” (see Table 2). Based on observation, this specific operation was difficult for technicians to perform and was error-prone. Technicians had to carefully plan a series of solution volumes to be pipetted. Such planning required as many as four chunks (coherent pieces) of information to be maintained in working memory (WM) at any given time, including: 1) a first solution volume; 2) the number of samples to be pipetted with the first solution; 3) a second solution volume; and; 4) the number of samples to be pipetted with the second solution. As a result of the store of other information in WM during this sequence, a total of eight chunks had to be attended by technicians until pipetting was complete. Previous research has suggested that cognitive overload and potential errors are likely to occur when more than five chunks of information must be maintained in WM (Kieras et. al, 1999; Lerch et al. 1989). Therefore, the sequence of recall and planning of the pipetting task was identified as a critical target for application of advanced robotic technology with the capacity to serve as an assistant to technicians and sharing in the cognitive task load as well as address follow-on perceptual-motor activities.

Table 4: Greatest numbers of sustained cognitive operations as part of long duration methods.

Methods	Max no. of cognitive sequence	Sequence description
Perform weighing of sample	9	Recall target weight, think of acceptable range of sample weight, and track current number of sample.
Tune device and prepare batch	7	Recall acceptable range of signal parameters, wait for stable signal, and think of how to adjust parameters.
Perform pipetting samples and solutions	13	Recall pipetting task parameters and plan for pipetting solution volume.

CONCLUSION

The objective of this research was to model life science lab technician performance in a complex chemical analysis process and to assess the cognitive workload imposed by various methods as bases for effectively directing process automation efforts. The findings of the study revealed technician perceptions of workload to be driven by method characteristics, including times, overall operation counts, and the number of motor and cognitive operations. Description of such characteristics, using cognitive task modeling methodologies, appears to be complementary to the use of subjective workload rating techniques as an approach for identifying and prioritizing tasks automation applications in order to reduce overall technician workload as well as the potential for task errors.

The cognitive task model developed in this research identified long duration methods, including sample weighing, tuning of analytical instruments and pipetting. These methods were also found to pose high workload for technicians based on subjective ratings. In general, the methods were considered to represent targets for application of process automation. The cognitive task models also revealed that among all process methods, weighing tuning and pipetting included the longest sequences of cognitive operations, which were identified by technicians as a major contributor to overall process workload. In specific, a sequence of task parameter recollections and planning of solution volumes for pipetting was found to represent the greatest sustained cognitive load on technicians. Consequently, this sequence of operations was identified as a priority target for use of advanced robotic technology with the capacity to act as assistants to lab technicians.

The use of the cognitive task modeling methodology combined with the cognitive workload assessment technique provided for a higher resolution analysis of the origin of cognitive load for technician and more precise specification of where potentially complex and costly process automation efforts should begin. The next step in this research is to perform a functional assessment of the available life science automation and robot technology for application to the chemical analysis process under study in order to achieve semi- and fully-autonomous operation for high-throughput of samples and high accuracy in analysis.

ACKNOWLEDGMENTS

This study was supported by the Federal Ministry of Education and Research (BMBF Germany; FKZ: 03Z1KN11) and CELISCA (University of Rostock, Germany) through a postdoctoral position for Manida Swangnetr. We thank the chemical process experts and lab technicians for critical information on the mercury analysis process. We also thank Weiping Zhang and Dr. Annika Rieger for configuration of the data collection devices used in the field study.

REFERENCES

Card, S., Moran, T., & Newell, A. (1983). *The psychology of human-computer interaction*. Erlbaum.

Diaper, D. (1989). *Task Analysis for Human-Computer Interaction* (Chap. 1, pp. 15-54). New York: Ellis Horwood Ltd.

Drury, J. L., Scholtz, J., & Kieras, D. (2007). Adapting goms to model human-robot interaction. In HRI '07: *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, (pp. 41-48), New York, NY, USA. ACM.

Endsley, M. R. (1993). A survey of situation awareness requirements in air-to-air combat fighters. *International Journal of Aviation Psychology*, *3*(2), 157–168.

Fleischer, H., & Thurow, K. (2013). Determination of Total Mercury Content in Wood Materials, Part 2: ICP-MS-A Multielement Method. *American Laboratory*, *45*(8), 6-8.

Foyle, D., Hooey, B., Byrne, M., Corker, K., Deutsch, S., Lebiere, C., Leiden, K., & Wickens, C. (2005). Human performance models of pilot behavior. In *Human Factors and Ergonomics Society Annual Meeting Proceedings*, volume 49, (pp. 1109-1113). Human Factors and Ergonomics Society.

Gray, W., John, B., & Atwood, M. (1992). The precis of project ernestine or an overview of a validation of goms. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, (pp. 307-312). ACM New York, NY, USA.

Haga, S., Shinoda, H., & Kokubun, M. (2002). Effects of task difficulty and time-on-task on mental workload. *Japanese Psychological Research*, *44* (3), 134-143.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: P. A. Hancock & N. Meshkati (eds.), *Human Mental Workload*. Amsterdam: North Holland Press.

Kaber, D. B., Wang, X. & Kim, S-H. (2011). Computational cognitive modeling of human-robot interaction using a GOMS methodology. In X. Wang (ed.), *Mixed Reality and Human Robot Interaction* (pp. 53-75). Dordrecht, The Netherlands: Springer.

Kieras, D.E. (2006). *A guide to GOMS model usability evaluation using GOMSL and GLEAN4* (On-line publication describing GOMSL, the GLEAN tool, and how to perform a GOMS analysis). University of Michigan, Department of Electrical & Computer Engineering, Ann Arbor, MI.

Kieras, D.E., Meyer, D.E., Mueller, S., & Seymour, T. (1999). Insights into working memory from the perspective of the EPIC architecture for modeling skilled perceptual-motor and cognitive human performance. In A. Miyake and P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. New York: Cambridge University Press. 183-223.

Lerch, F.J., Mantei, M.M., & Olson, J. R., (1989). Skilled financial planning: The cost of translating ideas into action. In *CHI'89 Conference Proceedings*, 121-126.

Polson, P. G. & Kieras, D. E. (1985). A quantitative model of the learning and performance of text editing knowledge. In *CHI '85: Proceedings of the SIGCHI conference on Human factors in computing systems*, (pp. 207-212), New York, NY, USA. ACM.

Salvucci, D. D. & Lee, F. J. (2003). Simple cognitive modeling in a complex cognitive architecture. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, (pp. 265-272), New York, NY, USA. ACM.

Swangnetr, M., Kaber, D. B., Vorberg, E., Fleischer, H., & Thurow, K. (current volume). Workload Assessment for Manual and Automated Processes in Life Sciences. In *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics AHFE 2014,* Kraków, Poland.

Swangnetr, M., Zhu, B., Kaber, D. B., Thurow, K., Stoll, N., & Stoll, R. (2012). Cognitive workload analysis and development of a stress index for life science processes. In *Proc. of the 2012 Applied Human Factors & Ergonomics Conference*. Boca Raton, FL: Taylor & Francis CRC Press.

Zhang, W., Stoll, R., Stoll, N., & Thurow, K. (2013). An mHealth monitoring system for telemedicine based on websocket wireless communication. *Journal of Networks*, *8*(4), 955-962.