Test Battery for Task Performance and Situational Awareness Measurement

Reza Karim
North Dakota State University
Reza.Karim@ndsu.edu

Kambiz Farahmand
North Dakota State University
kambiz.Farahmand@ndsu.edu

Abstract

In this study, a test instrument is developed that measure human cognitive task performance capacity, and situational awareness. Task capacity of a subject is measured before and after doing a low level of physical exercise. Signal Detection theory is applied to find the situational awareness of individuals in the form of true or false response. The tool compares the task capacity among different age and sex groups. The effect of low level stress on performance due to physical exercise is determined, and relationship is established between cognitive task and situational awareness. Response time, accuracy and repeatability are measured for statistical analysis. It compares pre and post stress effects on task performance. A stress level of the subject is measured using a questionnaire before participating on the test. The Delphi method is used to find the consensus on the test designed. The task capacity and situational awareness are expected to vary due to the effect of physical stress. Preliminary results indicate average accuracy of selecting correct answer is 84% and average time taken to complete the test 15.9 minutes. 81.4% of the task capacity related tasks and 89.9% of the situational awareness related tasks were correct in phase I of the experiment.

Introduction

Human perception, cognition and action takes place in a sequential manner; input is sensed, and then processed and output follows [1]. All current information on how the human mind works is based on this principle. Task capacity has been studied in many areas. Task capacity studies in Psychology or Clinical Psychology investigates relationships to mental disease. Neuroscience research evaluates the physical capability of the human brain. Human capacity can be classified into physical ability and task capacity. There are many ways to measure physical ability such as strength, oxygen consumption and heart rate. On the other hand situational awareness is the foundation of decision making. Task capacity is a complex process; there are many questions to answer on how to standardize and measure task capacity and situational awareness.

Rapid advancement of technology and increased complexity of work forces operators to adapt their decision making process in a dynamic environment. Workload is defined as the physical and mental requirements associated with a task. The proportion of physical load and mental load varies with the task types. Workload as a function of task requirement places demand on the human subject. The capacity of the subject is challenged to perform the workload. If the workload is higher than the operator capacity, the subject feels overloaded. Dynamic decision making process is considered as a real-time decision maker which is constrained by the decision making environment [2]. Plane and car accidents occur most often due to stresses on the operators of those vehicles. Stress, exceeding the tolerable limit, can be caused from physical conditions, physiological conditions or mental conditions at that time period. Environmental conditions and task loads develop stress not only in normal environmental conditions, but also in any type of confined environment. To find an appropriate person for a task, the assessment of task load and the impact it creates in performing the task is very important.

Background

Cognitive abilities play a key role in the adequate management of workload by individuals performing complex tasks [3]. Workload is dependent on task demand, and it varies on the capacity of the subject to meet those demands [4]. Studies show a decrement in performance as workload decreases [5]. A sudden decrease or increase of workload leads a loss of accuracy and slows response time in a longer work period. Experimental studies on cognitive performance usually keep workload constant.

Not much attention is paid in recent years on the relationship between task demand and cognitive abilities of the subject who performs the task. Some studies examined the effect of physiological activities on cognitive tasks. In the case of a short duration of physical activities, decrease of accuracy in performing cognitive tasks was observed, such as in the case of map interpretation [6]. Some findings indicate reverse results in arithmetic task performance at an intermediate level of physical task [7].

The need of a standard process describing and measuring task performance has well been recognized. Putting a right person on a right job is a tremendous challenge for companies. The tasks performed in any manufacturing facilities are routine and repetitive; efficiency and quality require standardized work procedures [8]. The standard process developed require being able to measure the effect of stress on cognitive task performance with greater degree of accuracy. Task performance capacity and physical performance capability vary from person to person [9]. These two conditions are independent of each other. For example a basketball player may be very good in physical activities but may not be as good in cognitive tasks. There are different mechanisms developed to measure physical abilities [10]. Not many methods are developed to measure cognitive task performance to find the stress effect. There is a challenge to develop an appropriate method and tool to simultaneously measure Cognitive Capability (CC) and Situational Awareness (SA) quickly and effectively.
Poor performance can occur using less friendly devices and could cause catastrophic error. Buckle [11] outlined the design challenges in the healthcare sector. The author provided

some approaches and methods that allow ergonomist to design any systems. Use of these design approaches was helpful to reduce the probability of medication error. Silver [12] studied the process carried out by the providers to improve the quality of the service provided based on the human factors approach. The key design considerations included task information characteristics, task allocation, redundancies and the competing goals of the operator. Spear [13] studied the ergonomics issues arising in the emergency department. The quality of healthcare service was improved by reducing the physical stress induced in the staff and patients due to the physical layout of the machines and equipments in a facility.

The Delphi method is generally used to develop the test procedure by consulting the experts in the specific field of application of the test battery. The test tasks and sequence of tasks are developed from expert opinion to fit the test objectives. The method standardizes the procedure to conduct the task capacity test. Goodman [14] compared the Delphi approach with committee decision approach. The Delphi method was chosen when the problem benefits from subjective statements made under a collective basis. Information was collected based on anonymity. Interpersonal interaction was eliminated to avoid the controlling variables in decision making. The key characteristics of the Delphi method, anonymity, use of experts and controlled feedback are examined in the Delphi study. Anonymity has advantages because it helps the participant to state their true opinion without being influenced by peer pressure. The disadvantage of the Delphi method is lack of accountability. On the other hand since the panel is selected on the basis of their knowledge and willingness to participate, the accountability problem may not be an issue. The validity of the study depends on the selection of experts instead of a random sample. Hasson [15] conducted methodological issues in nursing research, such as preparation, action steps and difficulties that are inherent within the Delphi technique. The validity of this method was enhanced by reasoned argument in which assumptions were challenged. Findings from the Delphi study helps streamline work. Three issues guide data collection: the discovery of opinion, the process of determining the most important issues and managing opinions-data analysis. The verbal approach, combined with the written approach, was found to be more effective in the The Delphi method. McKenna [16] described the Delphi technique and criteria for selecting it as a research tool. The Delphi method used for systematic collection and aggregation of information provided by the group of experts on specific questions and issues. The research population covers a diverse background in experience or expertise. If there is a lack of empirical data Delphi is appropriate. The unique aspect of this method is convergence towards agreement. It helps developing future knowledge and policy of a particular problem. Because of grassroots' involvement, the results from Delphi is widely accepted. Powell [17] emphasized the need on the development of scientific merit questions. Individual judgments recorded and combined in addressing the issues. The first round questionnaire is unstructured and obtains open response, allows participants to elaborate on the topic, and a qualitative analysis of the results allows constructing the second and subsequent questionnaires. The diversity of viewpoints that develop controversy help to generate interest and involvement. A heterogeneous group produces more high quality acceptable solutions than a homogeneous group. Villiers [18] described different types of Delphi technique: Conventional and real-time. In the conventional method, first a questionnaire is sent to a group of experts, and in the second round a questionnaire is sent back to the experts based on the result from the first

round. The third round is used depending on the consensus level from previous rounds. In the real-time technique the process takes place using a meeting where summary of the responses of the respondent is made immediately. The decision maker obtains information on options with supporting evidence from the forum and makes the decision. The forum does not make the decision.

Methodology

The test methodology developed here measures human cognitive task performance capacity and situational awareness simultaneously and compares task capacity among different age and sex groups. Cognitive task capacity and situational awareness of a subject is measured before and after doing a low level of physical exercise. Signal Detection theory is applied to find the cognitive task capacity and situational awareness of individuals in stress condition in the form of true or false response with measuring the confidence level of selecting the appropriate answer. Cognitive task capacity measure differs among the groups based on stress level of the subject. The task performance is measured in two stages. A low level physical work is set for the participant after the first stage. The tool determines if stress has any effect on the task performance. The test methodology determines any relationship between task performance and individual SA. Response time, accuracy, and repeatability of performing a given task are recorded for statistical analysis to justify the findings. A questionnaire developed through Delphi study is used to measure the level of stress level experienced by the subject participating in the test before hand. This pre-test creates a baseline of the candidate's stress level. If it is determined that the subject is stressed, the test will be rescheduled. The research focuses on determining the followings:

1. Develop the Task Capacity Model using performance parameters described by Miller [24]. Microsoft Visual C# 2.0 program in Microsoft.Net 2.0 platform to construct the model.
2. Standardize the Test using Delphi Techniques
3. Measure Task Capacity and Situational Awareness simultaneously using Signal Detection Theory
4. Determine dual task performance capacity
5. Objective queries and subjective self-ratings of confidence for each response determined
6. Determine any effect of low level stress on task performance

The focus of this research is to develop a standardized task capacity model. Human factor issues are considered to measure CC and SA simultaneously and to determine whether there is any effect of physical stress. Task complexity can be altered by changing the number of elements of a task. Task complexity effects attention, accuracy and repeatability of a task. The task capacity model build considers a standard task performance procedure created using the Delphi technique.

Mental capacity is the maximum potential to understand and follow the general logic of real world tasks from the user perception. Mental capacity combines two characteristics of the brain, one is the capacity to store and recall information (Memory capacity) and the other is the capacity to perform logic-processing operations (Problem-solving capacity). Problem-solving capacity and knowledge are independent measures of task capacity. But a high level of knowledge can enhance problem-solving efficiency. General Aptitude Test Battery (GATB) and the Employment National Job Service Committee (ENJSC) have been used in the United States for hiring people and to improve relationships between employers and employees. It has been described on the short comings of the GATB method by many authors [19]. Time given for the test is also a concern. IQ, SAT, ACT, academic records, GPA or work experience are considered for hiring people. Problem solving capacity and behavioral characteristics are not considered in many test methods developed.

Situational awareness is defined as the ability to identify the desired elements from the environment, process information and combine the critical elements of the information on the current situation. SA measures one's ability to recognize the present scenario and predict the future state of the gathered information. Performance parameter in a complex task model is dependent on SA. For example, in a flexible manufacturing system, operators must have up-to-date knowledge on machine tool parameters as well as the functioning for future process state changes [20]. Military personnel frequently rely on SA to make decisions in the battle field [21]. Inaccurate or incomplete SA could cause loss of life or unnecessary expenditure of resources. In recent years there are increased sophisticated military equipment used in the battle field which requires portable computing operations. The solders are required to be able to perform simultaneously cognitive demanding information processing tasks and physical tasks. Many studies show high SA score supports a good task performance. Stress may affect SA through decrement in working memory capacity and retrieval [22]. Literature suggests [23] that sensory tasks are enhanced by all levels of physical activities. It also finds improvement of memory and information processing by physical exercise.

Performance Parameters

Five performance parameters representing the real world tasks were described by Miller [24] as shown in Table 1. The twenty task functions are used to establish the relationship between task functions and task performance parameters. From the relationships between task functions and test parameters [24], the scoring technique is obtained. The equal brain capacity is assumed for each task.

Table1: Relationship of Task Functions and Task parameters

	Task Parameters				
Task Functions	Perception	Knowledge	Problem Solving	Memory	Creativity
	(PER)	(IQ)	(PSQ)	(MEM)	(CRE)
Message	X				
Input Select	X	X			
Detect	X		X		
Search/Locate	X			X	
Identify	X	X			
Filter		X	X	X	
Code			X	X	
Interpret		X			
Count			X	X	
Compute		X	X	X	
Decide/Select		X	X	X	
Compare			X	X	
Categorize		X	X	X	
Transmit		X		X	
Store			X	X	
Short-Memory				X	
Plan		X	X		X
Analyze		X	X	X	X
Adapt/Learn		X	X	X	X
Goal Image					X

Test Instrument

The test is designed to cover human task capacity and situational awareness in the following areas: computation, dual task, three-dimensional review, vocabulary, pattern recognition, comparison and arithmetic reasoning.

This study is broken into two phases, performed sequentially. The time gap between two tests is at aleast one week. Phase I will be conducted to determine task capacity and situational awareness simultaneously with a set of tasks in the form of questions. Phase II will compare stress produced by the physical activity performed right before participating in the experiment. There will be one experimental trial for each subject in Phase I. Each experimental trial consists of thirty tests in random order. Similarly Phase II will consist of one experimental trial with thirty tests in random order. Phase II will follow after fifteen minutes of light physical work at a set room air temperature and relative humidity. Subjects

will answer six stress level meaurement questions before and after the tests. Approximately sixteen subjects are desired to complete both the two phases of experiment. Ages range from twenty-one to forty. The selection process is random, anyone who is physically fit will be able to participate in the test. An individual apporoach will be conducted to recruit subjects.

This sample size provides a statistical power [25], 1-β, of 0.95 when using an analysis of variance to compare mean task capacity of at least eight individuals participating in the experiment assuming the study detects task capacity differences of 15% between phase I and phase II with a standard deviation of 7. The goal is to have a balanced experimental design for subsequent statistical analysis. It is desired that the same eight subjects participate in both experimental phases. However, if a subject do not continue after completion of phase I, they will not be replaced by other participants during phase II.

Test scenario is constructed using Delphi techniques. Test specifications used are: Measure of Respond Time (RT), Accuracy (AC) and Repeatability (RP). Cognitive capacity is measured in terms of IQ, MEM and PSQ. Situational Awareness is measured by describing a situation and after a set time period, situational related questions and scenarios are presented to the subject.

Human factor issues are considered when designing the test setup. The test is computer generated. Computer table and chair are positioned to allow the participant to adjust height to his or her comfort level.

All personal information is stored in the database with a unique user identification number. This is necessary because the same participant is expected to appear in the phase II experiment. The total number of participants considered for the test is ten. The same participants appear on Phase I and Phase II. Phase II is conducted after the participants perform a physical task for a specified amount of time to simulate stress.

This section describes the tool developed for the data collection. A Battery Test link is put on the Desktop to enter in to the program. A login information is provided to the subject to enter to the site.

Personal information is recorded in the first section. Part of the tool is shown in Figure 1.

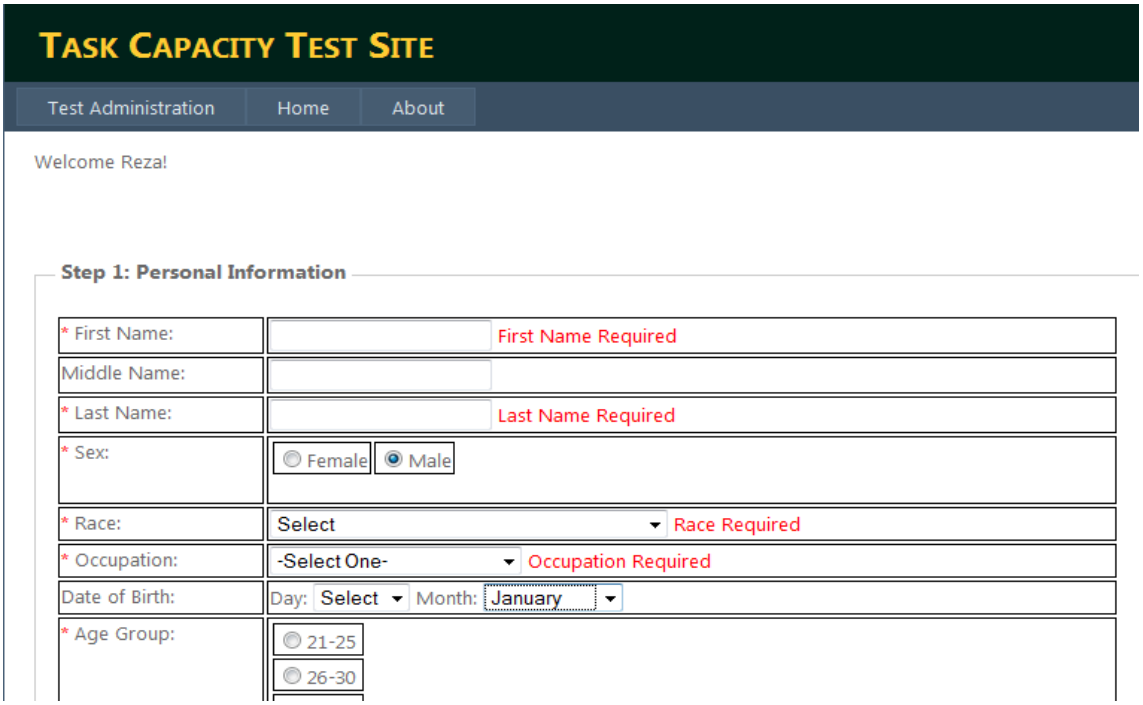


Figure 1: Snapshot of the personal information collection tool

After completion of the personal information, a stress level determination question is asked here.

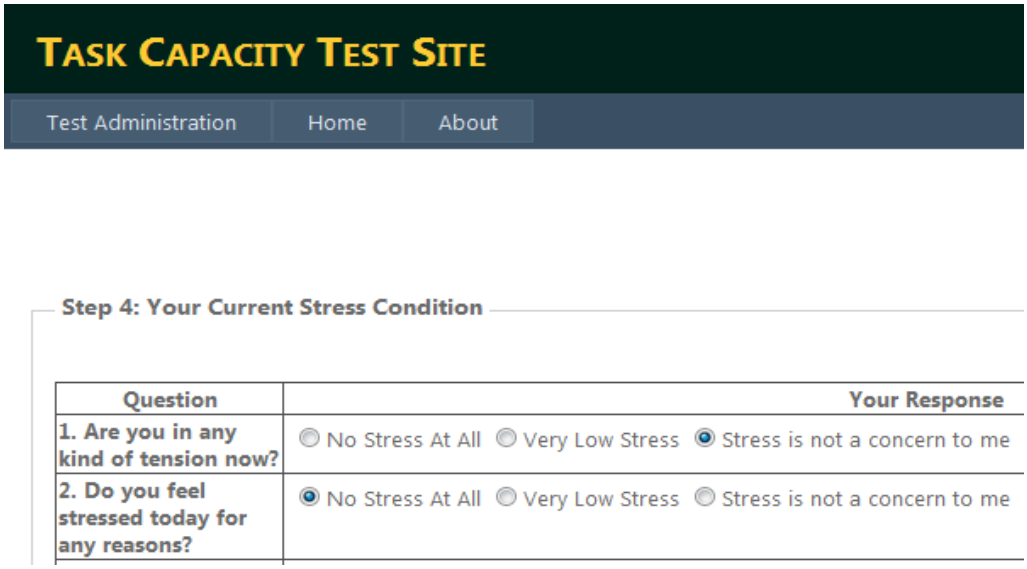
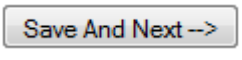


Figure 2: Snapshot of the Stress measurement tool

After the STRESS input the test will start by pressing 

The approximate time of the test is fifteen minutes. The following section provides an example of how the test scenario is designed.

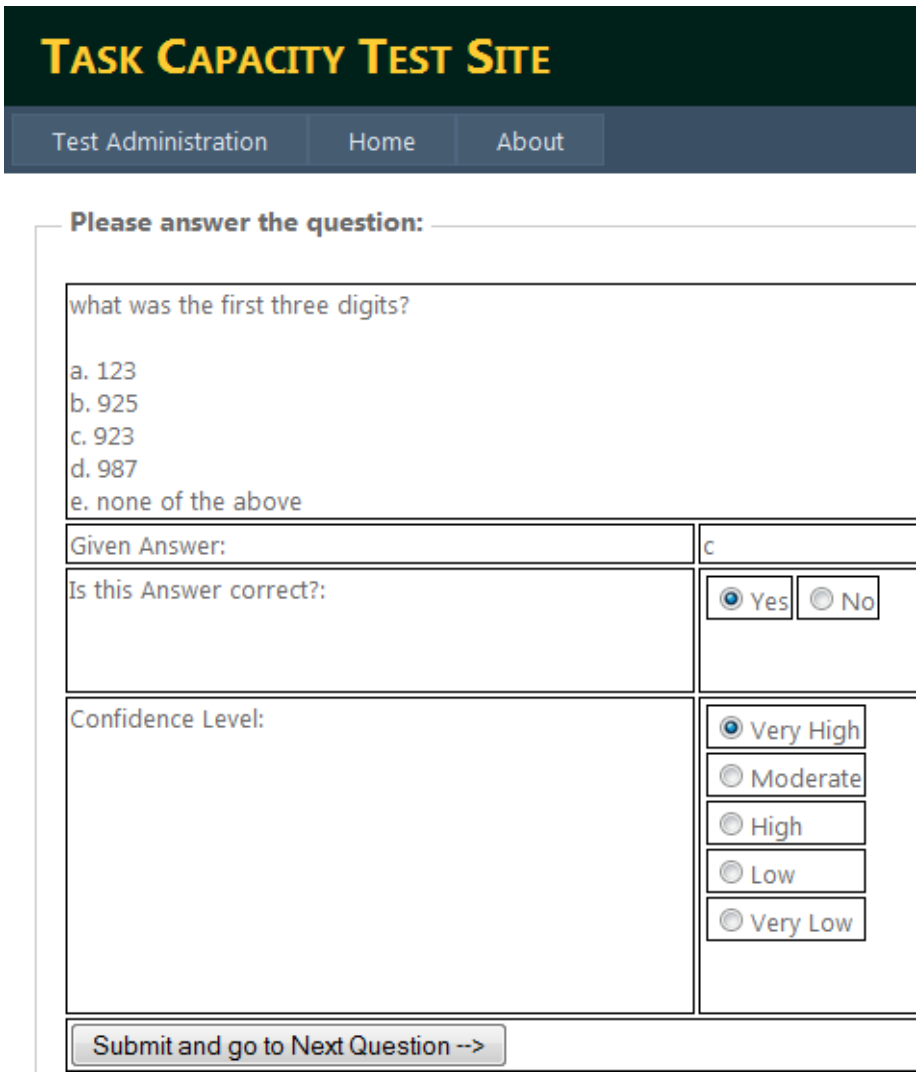


Figure 3: Snapshot of a questionnaire tool

The subject checks the appropriate button by comparing possible answers with the given answer. Right after selecting YES/NO, the subject selects a confidence level that describes how confident the subject is on selecting the answer. After completion of each question the subject presses the submit button and goes to the next level.

After a successful completion of the test STRESS level measurement question will appear. The subject needs to select the appropriate level that describes the subject's stress level.

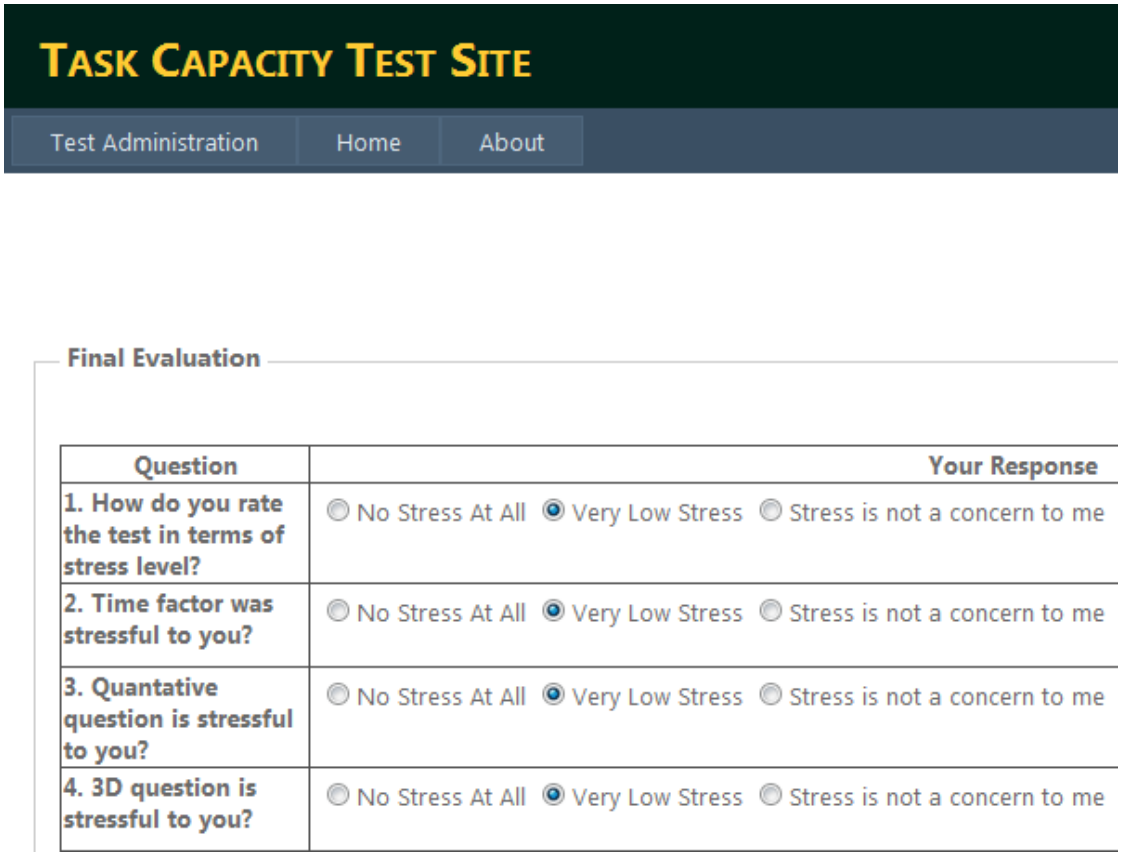


Figure 4: Snapshot of stress measurement tool

After finishing the stress test the subject presses Save-and-Finish Test button to complete the test.

Discussion & Result

Sample data is collected to verify the tool. Eight people were invited to participate in the phase I test, and five responses were obtained. Table 2 shows a sample of data collected for a subject.

Table 2: A table representing a portion of the data collected for each subject

Questions	Correct Ans.	Given Ans.	User Ans.	Is User Correct	Confidence Level	Response Time(ms)	Movement Time(ms)
1. Which pair of name is the same?	A	B	No	Yes	Very High	30265	3022
2. Add (+): 766 and 11	E	A	No	Yes	Very High	16786	1617
3. Which picture displays flat piece bent, rolled or both?	B	B	No	No	Very High	28284	1274
4. Which two words have the same meaning?	B & C	A & B	No	Yes	Very High	18658	1536

Table 3 summarizes response time, cursor movement time and accuracy of the responses for each subject.

Table 3: Response time and accuracy

	Phase I		
ID	**Response Time (RT) in seconds**	**Cursor Movement Time (RT) in seconds**	**Accuracy (AC) (%)**
01	419.79	366.70	85
02	1018.13	434.42	90
03	1504.24	679.99	80
04	918.02	116.56	90
05	915.04	535.47	75

Table 4 represents task capacity and situational awareness measured in percentage. There are forteen task capacity tasks and six situational awareness measurement tasks considered for the test.

Table 4: Task capacity and situation awareness

Phase I		
ID	**Task Capacity (%)**	**Situational Awareness (%)**
01	78.6	100
02	85.7	100
03	71.4	100
04	92.86	83.33
05	78.57	66.66

It is observed from the data that subjects make incorrect selections even at high confidence level. The participants made 89.9% correct selections for the situational awareness tasks.

Conclusion

The current research focuses on standardizing the task functions to measure individual task capacity. Delphi method which is usually applied in social policy and public health is considered as a research tool to determine the needs and skills require in any specific work environment. Using a single tool simultaneously to measure cognitive task capacity and situational awareness is expected to be a useful application in the dynamic and complex work environment of manufacturing industries. The low level stress becomes a challenge on cognitive task performance when repetitive tasks are performed. The method developed in this research is expected to differentiate the type of task functions that are affected significantly when stress is a concern. Self-rated stress measurement examines what type of tasks may be considered as a stressor to individuals.

References

[1] Kroemer, Karl, Kroemer, Henrike and Kroemer, Katrin, "Ergonomics: How to Design for Ease and Efficiency," Prentice Hall, 2nd Edition, 2003.

[2] Edwards, W., "Dynamic decision theory and probabilistic information processing," Human Factors, 4, 1962, pp 59-73.

[3] Gonzalez, Cleotilde, "Task Workload and Cognitive Abilities in Dynamic Decision," Human Factors**,** Vol. 47, No. 1, 2005, pp. 1-10.

[4] Gopher, D. and Donchin, E., "Workload – An examination of the concept," Handbook of Perception and Human Performance. Volume 2. Cognitive Processes and Performance. K.R. Boff, L. Kaufman and J.P. Thomas, John Wiley and Sons, Inc: 41-1:41-49, 1986.

[5] Cox-Fuenzalida,Luz-Eugenia, 2007. Human Factors**,** Vol. 49, No. 2, pp. 277-291.
[6] Hancock, S. and McNaughton, L., "Effects of fatigue on ability to process visual information by experienced orienteers," Perceptual & Motor Skills, 62, 1986, pp. 491-498
[7] Reilly, T. and Smith, D., "Effect of work intensity on performance in a psychomotor task during exercise," Ergonomics, 29, 1986, pp. 601-606.
[8] Adler, Paul S., "Time-and-Motion Regained,' Harvard Business Review, 2000.
[9] Weimer, Jon, "Research techniques in human engineering,", Englewood Cliffs, N.J: Prentice Hall, 1995.
[10] Sanders, Mark S. and McCormick, Ernest J., "Human Factors in Engineering and Design," McGraw-Hill, 7th edition, 1993.
[11] Buckle, P. et al., "Patient safety, systems design and ergonomics," Applied Ergonomics 37, 2006, pp 491-500.
[12] Silver, M.P. et al., "Improving Healthcare systems performance: A Human Factors approach," American journal of medical quality 19(3), 2004.
[13] Spear, M.E., "Ergonomics and Human Factors in healthcare settings,' Annals of emergency medicine 40(2), 2002, pp 213-216.
[14] Goodman, Ckaire et al., "The Delphi Technique: a critique," Journal of Advanced Nursing 12, 1987, pp 729-734.
[15] Hasson, Felicity et al., "Research Guideline for the Delphi Survey technique," Journal of Advanced Nursing 32 (4), 2000, pp 1008-1015.
[16] McKenna, Hugh et al., "The Delphi Technique: a worthwhile research approach for nursing," Journal of Advanced Nursing 19, 1994, pp 1221-1225.
[17] Powell, Catherine, "The Delphi technique: myths and realities," Journal of Advanced Nursing 41(4), 200, pp 376-382.
[18] Villiers, Mariet et al., "The Delphi Technique in Health Sciences Education Research," The Journal of Psychology 113, 1987, pp 73-80.
[19] Hartigan, John A., and Wigdor, Alexandra K., "Fairness in Employment Testing: Validity Generalization, Minority Issues, and the General Aptitude Test Battery," National Academy Press, Washington, D.C., 1989, pp 113-116.
[20] USHER, J.M. and KABER, D.B., "Establishing information requirements for supervisory controllers in a flexible manufacturing system using GTA," Human Factors and Ergonomics in Manufacturing, 10, 2000, pp 431–452.
[21] KABER, D.B., RILEY, J.M., LAMPTON, D. and ENDSLEY, M.R., "Measuring situation awareness in a virtual urban environment for dismounted infantry training," In Proceedings of the 1st International Conference on Virtual Reality (CD-ROM), Las Vegas, NV, 22–27 July, 2005.
[22] Endsley, Mica R., 'Direct Measurement of Situational Awareness in Simulations of Dynamic Systems: Validity and use of SAGAT," Presented in the International Conference on Experimental Analysis and Measurement of Situational Awareness, Daytona Beach, Florida, 1995.
[23] Tomporowski, P., & Ellis, N., "Effects of exercise on cognitive processes: A review," Psychological Bulletin, 99(3), 1986, pp 338–346.

[24] Miller, Robert B., "A Method for Determining Task Strategies," American Institutes for Research in the Behavioral Sciences, Silver Spring, MD. Report number APHRL-TR-74-26, 1974.
[25] Montgomery, Douglas C., "Introduction to Statistical Quality Control," Wiley, 4th Edition, 2001

Biographies

REZA KARIM is in Ph.D. program at Industrial Engineering Department at North Dakota State University. He is currently conducting research in improving Healthcare facilities design and management. His research interest is in the area of healthcare, simulation, product design, process analysis, lean application and human factor consideration. He has work experience in the field of equipment design, process development, CFD analysis and project management.

KAMBIZ FARAHMAND is currently a Professor at the Industrial and Manufacturing Engineering and Management at North Dakota State University. He is an internationally recognized expert in Productivity Improvement. Dr. Farahmand has over 28 years of experience as an engineer, manager, and educator. He is a registered professional engineer in the state of Texas and North Dakota.