TEST BATTERY FOR EVALUATION OF TASK PERFORMANCE AND SITUATIONAL AWARENESS

Reza Karim, North Dakota State University; Kambiz Farahmand, North Dakota State University

Abstract

In this study, a test instrument was developed to measure human cognitive task-performance capacity and situational awareness. Task capacity of a subject was measured before and after doing a low level of physical exercise. Signal-detection theory was applied to find the situational awareness of individuals in the form of true-or-false responses. The tool compares the task capacity among different age and sex groups. The effect of low-level stress on performance due to physical exercise was determined, and a relationship was established between cognitive task and situational awareness. Response Time, accuracy and repeatability were measured for statistical analysis. It compared pre- and post-stress effects on task performance. A stress level of the subject was measured using a questionnaire before participating on the test. The Delphi method was used to find the consensus on the test designed. The task capacity and situational awareness were expected to vary due to the effect of physical stress. Preliminary results indicated that average accuracy of selecting the correct answer was 84% and the average time taken to complete the test was 15.9 minutes. Also, 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, followed by an output [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 investigated 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.

The rapid advancement of technology and the 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 demands 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. A 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 by physical conditions, physiological conditions or mental conditions. 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 varies with 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 to 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 has been paid in recent years to the relationship between task demand and cognitive abilities of the subject performing the task. Some studies examined the effect of physiological activities on cognitive tasks. In the case of a short duration of physical activities, a 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 the right person on the job is a tremendous challenge for companies.

The tasks performed in any manufacturing facility are routine and repetitive; efficiency and quality require standardized work procedures [8]. The standard process requires 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 ergonomists 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 equipment 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 the 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, were examined in the Delphi study. Anonymity has advantages because it helps the participants 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 help 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 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 related to the subject of concern.

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 in developing knowledge and policy of a particular problem. Because of grassroots' involvement, the results from Delphi activities are widely accepted. Powell [17] emphasized the need for the development of scientific merit questions. Individual judgments recorded and combined in addressing the issues. The first-round questionnaire was unstructured and obtains open responses, 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 develops controversy helps to generate interest and involvement among the selected experts. A heterogeneous group produces more high-quality acceptable solutions than a homogeneous group. Villiers [18] described different types of Delphi techniques: conventional and real time. In the conventional method, a questionnaire is first sent to a group of experts; 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 a 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 stressful conditions in the form of true-or-false responses while measuring the confidence level of selecting the appropriate answer. Cognitive task-capacity measures differ among the groups based on stress level of the subject. Task performance is measured in two stages. 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 a Delphi study was used to measure the level of stress 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 focused on determining the following:

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 was 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 affects attention, accuracy and repeatability of a task. The task-capacity model considers a standard task-performance procedure created using the Delphi technique.

Mental capacity is the potential to understand and follow the general logic of real-world tasks from the user's 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. The 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. The GATB has been described as having shortcomings 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.

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 parameters in a complex task model are 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 on the battle field [21]. Inaccurate or incomplete SA could cause loss of life or unnecessary expenditure of resources. In recent years, the military has employed increasingly sophisticated equipment on the battle field, which requires portable computing operations. The solders are required to be able to perform simultaneous, cognitively demanding information-processing tasks and physical tasks. Many studies show that high SA scores support good task performance. Stress may affect SA through a decrease in working memory capacity and retrieval [22]. A preview of the literature suggests [23] that sensory tasks are enhanced by all levels of physical activities. It can also be seen that 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. Equal brain capacity is assumed for each task.

Table 1. Relationship of Task Functions and Task Parameters

Task Functions	Task Parameters				
	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

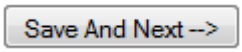
The test was designed to cover human task capacity and situational awareness in the following areas: computation, dual task, three-dimensional review, vocabulary, pattern recognition, comparison and arithmetic reasoning.

This study was broken into two phases, performed sequentially. The time gap between two tests is at least one week. Phase I was conducted to determine task capacity and situational awareness simultaneously with a set of tasks in the form of questions. Phase II compared stress produced by the physical activity performed right before participating in the experiment. There was one experimental trial for each subject in Phase I. Each experimental trial consisted of thirty tests in random order. Similarly, Phase II consisted of one experimental trial with thirty tests in random order. Phase II followed after fifteen minutes of light physical work at a set room air temperature and relative humidity. Subjects answered six stress-level measurement questions before and after the tests. Approximately sixteen subjects were desired to complete the two phases of the experiment. Ages ranged from twenty-one to forty. The selection process was random; anyone who was physically fit could participate in the test. An individual approach was 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 was to have a balanced experimental design for subsequent statistical analysis. It was desired that the same eight subjects participate in both experimental phases. However, if subjects could not continue after completion of Phase I, they would not be replaced by other participants during Phase II.

The test scenario was constructed using Delphi techniques. Test specifications used were: Measure of Response Time (RT), Accuracy (AC) and Repeatability (RP). Cognitive capacity was measured in terms of IQ, MEM and PSQ. Situational Awareness (SA) was measured by describing a situation and after a set time period, situational-related questions and scenarios were presented to the subject. Human-factor issues were considered when designing the test setup. The test was computer generated. Computer tables and chairs were positioned to allow the participants to adjust the height to their comfort level. All personal information was stored in the database with a unique user-identification number. This was necessary because the same participant was expected to appear in Phase II of the experiment. The total number of participants considered for the test was ten. The same participants appeared in Phase I and Phase II. Phase II was conducted after the participants performed a physical task for a specified amount of time to simulate stress.

This section describes the tool developed for data collection. A Battery-test link was put on the desktop to enter into the program. Login information was provided to the subject to enter onto the site. Personal information was recorded in the first section. Part of the tool is shown in Figure 1. After completion of the personal information, a stress-level determination question was asked (see Figure 2). After the STRESS input, the test would start by pressing



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

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 he/she is on selecting the answer. After completion of each question, the subject presses the submit button and goes to the next level.

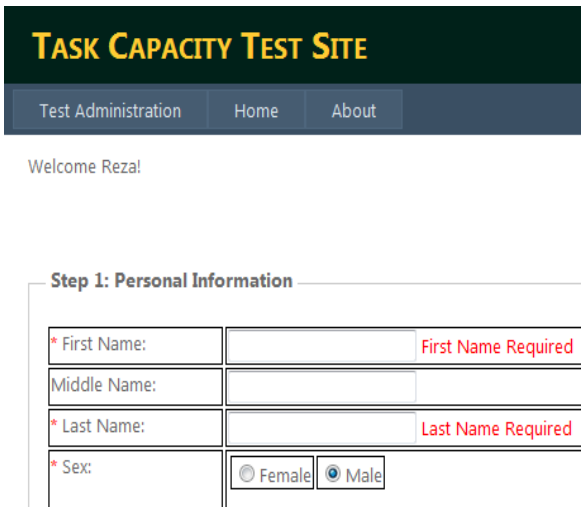


Figure 1. Snapshot of the Personal Information Collection Tool

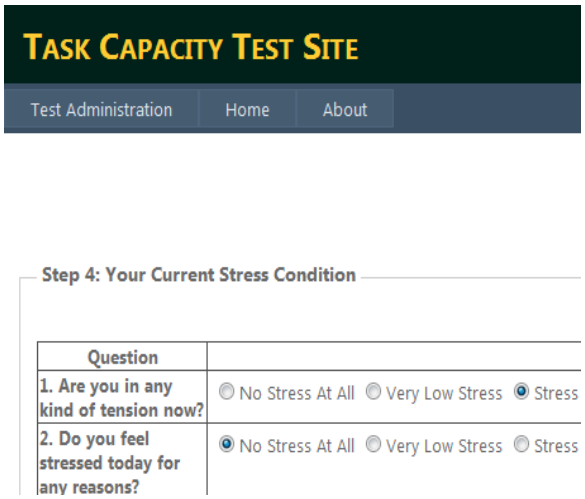


Figure 2. Snapshot of the Stress Measurement Tool

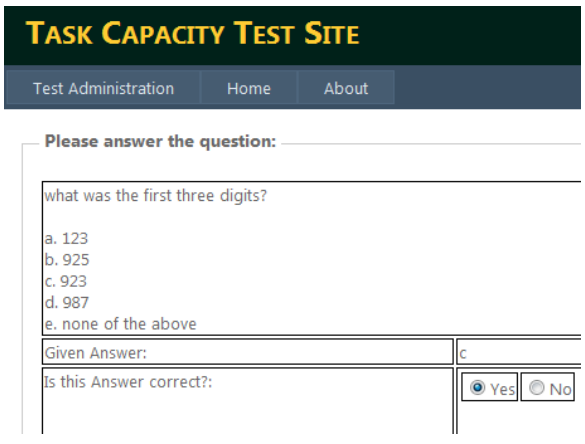


Figure 3. Snapshot of a Questionnaire Tool

After a successful completion of the test, a STRESS-level measurement question would appear. The subject needed to select the appropriate level that described his/her stress level. After finishing the stress test, the subject presses the Save-and-Finish Test button to complete the test.

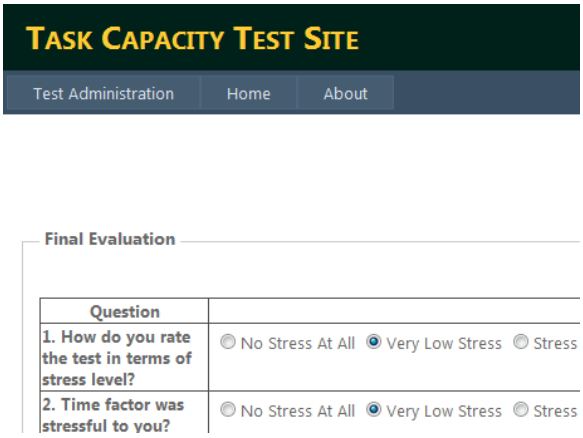


Figure 4. Snapshot of Stress Measurement Tool

Discussion & Results

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

Table 2. A Sample Representation 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

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 fourteen 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 was observed from the data that subjects make incorrect selections even at high-confidence levels. The participants made 89.9% correct selections for the situational-awareness tasks.

Conclusion

The current research focused on standardizing the task functions to measure individual task capacity. The Delphi method, which is usually applied in social policy and public health, was considered as a research tool to determine the needs and skills required in any specific work environment. Using a single tool simultaneously to measure cognitive task capacity and situational awareness was expected to be a useful application in the dynamic and complex work environment of manufacturing industries. 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, K., Kroemer, H. & Kroemer, K. (2003). *Ergonomics: How to Design for Ease and Efficiency.* (2nd ed.). Prentice Hall.

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

[3] Gonzalez, C. (2005). Task Workload and Cognitive Abilities in Dynamic Decision. *Human Factors***,** 47 (1), 1-10.

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

[5] Cox-Fuenzalida, L-E. (2007). Effect of Workload History on Task Performance. *Human Factors***,** 49, (2), 277-291.

[6] Hancock, S. & McNaughton, L. (1986). Effects of fatigue on ability to process visual information by experienced orienteer. *Perceptual & Motor Skills*, 62, 491-498.

[7] Reilly, T. & Smith, D. (1986). Effect of work intensity on performance in a psychomotor task during exercise. *Ergonomics*, 29, 601-606.

[8] Adler, P. S. (2000). Time-and-Motion Regained. *Harvard Business Review*.

[9] Weimer, J. (1995). *Research techniques in human engineering***.** Englewood Cliffs, N.J: Prentice Hall.

[10] Sanders, M. S. & McCormick, E. J. (1993). *Human Factors in Engineering and Design*. (7th ed.). McGraw-Hill.

[11] Buckle P. , Clarkson, P. J., Coleman, R., Ward, J. & Anderson, J. (2006). Patient safety, systems design and ergonomics. *Applied Ergonomics*, 37, 491-500.

[12] Silver, M. P., Geis, M. S. & Bateman, K. A. (2004). Improving Healthcare systems performance: A Human Factors Approach. *American Journal of Medical Quality*, 19(3).

[13] Spear, M. E. (2002). Ergonomics and Human Factors in healthcare settings. *Annals of emergency medicine*, 40(2), 213-216.

[14] Goodman, C. (1987). The Delphi Technique: a critique. *Journal of Advanced Nursing*, 12(6), 729-734.

[15] Hasson, F., Keeney, S. & McKenna, H. (2000). Research Guideline for the Delphi Survey technique. *Journal of Advanced Nursing*, 32(4), 1008-1015.

[16] McKenna, H. (1994). The Delphi Technique: a worthwhile research approach for nursing. *Journal of Advanced Nursing*, 19(6), 1221-1225.

[17] Powell, C. (2003). The Delphi Technique: Myths and Realities. *Journal of Advanced Nursing*, 41(4), 376-382.

[18] Villiers, M. R., Villiers, P. J. & Kent, A. P. (1987). The Delphi Technique in Health Sciences Education Research. *The Journal of Psychology*, 113, 73-80.

[19] Hartigan, J. A. & Wigdor, A. K. (1989). *Fairness in Employment Testing: Validity Generalization, Minority Issues, and the General Aptitude Test Battery*. Washington, D.C.: National Academy Press.

[20] Usher, J. M. & Kaber, D. B. (2000). Establishing information requirements for supervisory controllers in a flexible manufacturing system using GTA. *Human Factors and Ergonomics in Manufacturing*, 10,

431–452.
[21] Kaber, D. B., Riley, J. M., Lampton, D. & Endsley, M. R. (2005, July). *Measuring situation awareness in a virtual urban environment for dismounted infantry training*. Proceedings of the 1st International Conference on Virtual Reality, Las Vegas, NV.
[22] Ensley, M. R. (1995). *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.
[23] Tomporowski, P., & Ellis, N. (1986). Effects of exercise on cognitive processes: A review. *Psychological Bulletin*, 99(3), 338–346.
[24] Miller, R. B. (1974). *A Method for Determining Task Strategies*. American Institutes for Research in the Behavioral Sciences, Silver Spring, MD. Report number APHRL-TR-74-26.
[25] Montgomery, D. C. (2001). *Introduction to Statistical Quality Control*. (4th ed.). Wiley.

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. Mr. Karim may be reached at Reza.Karim@ndsu.edu

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. Professor Farahmand may be reached at Kambiz.Farahmand@ndsu.edu