

# Cyber Operations Stress Survey (COSS): Studying fatigue, frustration, and cognitive workload in cybersecurity operations

Josiah Dykstra  
U.S. Department of Defense  
jdykstra@LTSnet.net

Celeste Lyn Paul  
U.S. Department of Defense  
clpaul@tycho.ncsc.mil

## Abstract

Operator stress is a common, persistent, and disabling effect of cyber operations and an important risk factor for performance, safety, and employee burnout. We designed the Cyber Operations Stress Survey (COSS) as a low-cost method for studying fatigue, frustration, and cognitive workload in real-time tactical cyber operations. The combination of pre- and post-operational measures with well validated factors from the NASA Task Load Index and additional contextual factors provide a quick, easy, and valuable assessment of cognitive stress. We report on our experiences developing and fielding the survey instrument, validation, and describe the use and results of the COSS in four studies of cyber operations across the National Security Agency.

## 1 Introduction

Cybersecurity is a high-risk, high-reward profession that can negatively impact a company's technical workforce. While considerable research has helped evaluate and improve technology resiliency, *human* resiliency has been understudied despite the important role of humans in the design and execution of cybersecurity programs [4]. In this paper, we focus on a complimentary goal of measuring human distress which can severely impact operational effectiveness and human health. In particular, we offer a new research instrument for measuring and assessing stress in tactical cyber operations.

Over the past decade, cybersecurity operations have greatly matured. Security monitoring in many organizational environments occurs internally and as a managed service. Security Operations Centers (SOCs) offer one example of this, where dedicated security teams perform threat monitoring, investigation, mitigation, and response to security events. Tasks in the SOC require vigilance of changing threats, increasing volume of alerts, and incomplete monitoring. Other than extraordinary

circumstances, such as the discovery of an attack in progress (e.g., distributed denial-of-service) or the discovery of a sensitive data breach, defensive operations typically lack significant time pressure.

**Tactical cyber operations.** We distinguish a subset of cyber operations called *tactical cyber operations*, in which cyber capabilities are used to achieve specific effects on a network. Capture the flag games for military exercises such as USCYBERCOM's annual Cyber Flag event are an example of this type of work [18]. Another example is red team penetration testing, where an independent group plays the adversarial role and 'attacks' an organization to test that organization's defenses.

Tactical cyber operations are unique in several respects. Performance is highly dependent on speed and precision, just as it is for fighter pilots and surgeons. The longer operation, the greater the risk, such as increased likelihood of unintended detection on the network. Tactical operators require specialized skills and traits. For examples, penetration testers have a breadth of expertise in network and software fundamentals, reconnaissance, exploitation, and adversarial thinking. Training for this type of work is extensive, expensive, and employee turnover is costly. The health of your talent is as much of a risk management issue as it is a human resources issue.

**Why we care about stress.** A key motivation for this work is the intuition that stress negatively affects operational security, work performance, and employee satisfaction. Tasks that involve attention, memory, and visual perception result in high levels of cognitive demand and fatigue. There is a strong connection between fatigue and stress [21], and fatigue and task performance [12]. We know that stress negatively affects cognitive abilities, task effectiveness, and general well-being. These types of effects are harmful to high-risk, mission-critical environments where failure has great consequence. Stress is detrimental to work that requires creative problem solving—a skill that cyber operators inherently require.

## 2 Stress and Cyber

*Stress* is a physical and emotional reaction to certain types of situations. It is very subjective and everyone experiences it differently. *Acute stress* is the most common form of stress resulting in a ‘fight or flight’ response; however, once the stressor goes away, so do the symptoms. *Episodic stress* is when acute stress occurs frequently and you do not have enough time to recover from stress event to event, resulting in lower tolerance and higher sensitivity to stress events. *Chronic stress* is long-term stress results from enduring situations where you lack control over the outcome and can have serious effects on physical and mental health. Episodic and chronic stress contribute to increased operational risk and employee burnout.

Many factors can affect or be affected by stress. Our work focuses on factors that tend to have the biggest impacts on work. *Fatigue* is the feeling of tiredness that may result in the temporary inability to maintain optimal performance in a mental or physical capacity. *Frustration* is the anxiety and annoyance you experience in response to uncertainty and insecurity that stems from a sense of inability to fulfill needs, goals, or desires. *Cognitive workload* is the amount of mental effort needed to utilize and execute working memory. Measures of cognitive workload can characterize the attentional demands that a task places on an individual.

**Stress and work.** Stress, error, and burnout have been studied extensively in many non-cyber domains. Grier found that across several areas of research, the most demanding cognitive domains were air traffic control, command and control, and medical tasks [7]. Each of these tasks high amounts of attention, memory, and visual perception. Workload is important to study and understand because high workload increases the likelihood of errors and decreases performance [14]. In a study of individuals across professional backgrounds, burnout was closely related to mental exhaustion due to stress [21]. In a study of emergency management, Kowalski-Trakofler and Vaught found that judgment is not always compromised under stress, but that the effects of stress on judgment and cognition is highly contextual and cannot be easily generalized [13].

**Stress in cyber.** Stress and cognitive workload among network defenders and analysts is drawing attention from a wide community of researchers and cyber defense organizations. Work by Greenlee et al. found increases in cognitive workload and stress going from initial network intrusion triage to escalation analysis [6]. The U.S. Air Force found that shift work, shift changes, and hours worked contributed to high occupational stress and burnout in cyber warfare operators [3]. In a 40-minute simulated cyber security experiment, Sawyer et al. found

that the required vigilance for cyber events was consistent with results from air traffic control, industrial process control, and medical monitoring and that mental workload was considerably high [20]. A 2016 study of cybersecurity incident responders revealed sources of stress from the urgency to respond to a cyber event and the hyper vigilance involved from not wanting to reveal position or strategy to network intruders who could still present in the network [8].

**Challenges of human research in cyber.** Despite growing efforts to understand the intersection between people and cyber, challenges of doing research in cybersecurity remain. The population of security practitioners, while growing, is still relatively small. They are also overworked, stressed, and generally inaccessible or unable to participate in research due to corporate privacy requirements. Additionally, methods in human subjects research is expensive, time consuming, and difficult to execute. The field of human research in cyber is also still immature and there are few validated research methods and tools tailored to this environment.

Network and security operations centers are often sensitive information environments and closely guarded and generally unavailable to even those within the same company [15]. Data collected and analyzed may contain proprietary or personally identifiable information. Intrusions and vulnerabilities are often closely held secrets. Gaining access to these operational details by researchers can be challenging despite being valuable information for understanding the cyber operation environment.

Recruiting study participants is always difficult in cybersecurity research [1], and we experienced similar challenges. There are a limited number of tactical cyber operators in the environment, and time in a study meant time away from critical mission activities. This population also suffered from ‘survey fatigue’, having been studied repeatedly for other purposes.

## 3 Cyber Operations Stress Survey (COSS)

In an effort to improve the state of research in human-focused cybersecurity research, we present the Cyber Operations Stress Survey (COSS). We developed the COSS over the course of four independent studies of tactical cyber operations. While the results of those studies are described in [5, 16, 17], this paper focuses on the details of the COSS methodology so that it can be utilized by more cybersecurity researchers.

The COSS is a subjective, summative assessment of an operator’s stress during a tactical cyber operation. As previously discussed, stress can be characterized in many ways. The factors we care about most are those that affect performance and creative problem-solving abilities: fatigue, frustration, and cognitive workload.

The COSS is influenced by experience sampling methods [11] and was designed as a study instrument that does not take a lot of time and can be used for repeated sampling. Our approach combines several well-known stress assessments to provide a well-validated, easy to execute instrument for studying operations in the field. One of our design goals was minimally-invasive, in situ reporting, and we did not use methods such as the Job-related Affective Well-being Scale (JAWS) [22] because of the length of the assessment.

**Samn-Perelli Fatigue Scale (SPFS).** The SPFS was initially developed by the U.S. Air Force to measure pre-flight fatigue in pilots [19]. The question asks study subjects to indicate: “How awake, or tired are you?” The scale can be used anchored or unanchored. The COSS uses a 20-point anchored scale that qualifies different fatigue states, to help normalize subjective feelings of tiredness across sessions and between participants.

**NASA Task Load Index (TLX).** The NASA TLX is a popular subjective assessment use in engineering to measure cognitive workload along six dimensions on a 20-point unanchored scale [9]. These dimensions are mental demand, physical demand, time demand, subjective performance, frustration, and effort. The TLX is a well validated instrument and has been used in hundreds of studies of information technology. Many items in longer stress surveys, such as JAWS, are included in the TLX.

**Fatigue and Frustration Baseline.** Some types of stress are more sensitive to individual differences than others and can be influenced by external factors independent of a study. For example, how tired someone is at the start of an operation can vary greatly, while the amount of effort to complete an operation is more likely to be dependent on the operation. Pilot studies of the COSS showed that the nature of the tactical cyber operations environment led to high levels of external frustration that sometimes interacted with operational frustration. For this reason, the COSS baselines pre-operation fatigue and frustration to compare with post-operation differences and to provide a method to normalize differences across sessions and between participants.

**Other Contextual Measures.** The operational task or activity may not be the only source of cognitive stress. External factors related to personal life (e.g., illness), physical environment (e.g., noise level), or even other work duties (e.g., an upcoming deadline) may cause or interact with cognitive stressors. Depending on the study environment, information about these types of factors may provide useful and necessary context for interpreting results of the COSS. Certain types of demographics (e.g., years on the job) may also be relevant to a study.

**Study Execution.** The first part of the survey (pre-operation) is administered at the start of the operation and the second part of the survey (post-operation) is ad-

ministered immediately at the end of the operation. The survey can be used in a repeated-measures study (as is often the case in user experience reporting studies). A minimized demographics questionnaire is built into the survey so that new participants can provide their data, as needed. The survey may be executed as a paper-based or web-based survey. The paper-based survey can be useful because the paper acts as a physical reminder to complete the post-operation section. However, some participants may prefer to complete a web-based form. Data collection and analysis is easier with a web-based survey, especially in large-scale studies.

**Data Analysis.** Pre- and post-operation fatigue and frustration measures can be analyzed in two ways. First, the differences between pre- and post-operation fatigue and frustration can be measured using a paired means or rank test. Second, these measures can be converted into a  $\Delta$ Fatigue and  $\Delta$ Frustration metric for use as a descriptive statistic or comparison between groups. The TLX can be analyzed in several ways. The traditional method calculates an overall weighted metric based on the six workload measurements [9], while a more modern approach reports an unweighted metric (RTLX) [9, 17]. Another common analysis is to report each individual workload measure to preserve context. The TLX does not have a ‘redline’ test in which it can be used as a quantitative threshold of what could be considered ‘too high’ of a workload [9]. We recommend the RTLX score and individual workload measures for analysis of the COSS. Contextual factors should be analyzed as appropriate. For example, some demographic information is often nominal and can be used to test differences between groups while our example of a ‘team synergy’ measure could be analyzed as an ordinal or interval.

## 4 Case Study

As proof of concept for this method, we describe how we used COSS in four independent studies of tactical cyber operations at the National Security Agency [5, 16, 17]. The NSA coordinates, directs, and performs highly specialized activities to protect U.S. government information systems and to produce foreign signals intelligence. In addition to its headquarters in Maryland, NSA has cryptologic centers in Colorado, Georgia, Hawaii, and Texas that also conduct foreign signals intelligence, cyberspace operations, and information assurance operations. Tactical cyber operators at these locations have competencies in operating systems, network analysis, network penetration testing, intrusion detection, incident response, digital forensics, as well as strong critical thinking and problem solving skills. New operators complete up to six months worth of training and 85% of operators have at least two years of experience.

In addition to the base COSS questions (SPFS, TLX, pre- post- baseline), we included additional contextual measures specific to our environment (not all are listed in the sample instrument in the Appendix). For example, our operators work irregular schedules, and so we asked what time they came in to the office to judge how much office time they had before an operation, and when they last had a day off. They also often work in a team environment, and so we included a measure of ‘team synergy’ between the participant and his teammates.

## 4.1 Study Results

Across four studies, 126 total participants completed 361 surveys (Table 1). Participation was voluntary, and subjects could complete one survey for any operation during the study period. The average length of an operation was 5.12 hours (SD = 2.0 hours).

Study	Participants	Surveys	Length	Mode
Site 1	32	56	5 weeks	Paper
Site 2	31	102	5 weeks	Paper
Site 3	23	67	3 weeks	Web
Site 4	40	136	3 weeks	Web
<i>Total</i>	<i>126</i>	<i>361</i>	<i>3-5 weeks</i>	<i>Both</i>

Table 1: Summary of four studies using the COSS

A more detailed report of the results of these studies is reported in [17]. We summarize these results to provide context for the COSS survey method described in Section 3 and instrument validation discussion in Section 4.2.

**Fatigue and Frustration.** Operator fatigue and frustration increased significantly over the course of an operation (Figure 1). Fatigue increased by 16%, *Student’s*  $t(359) = -13.92, p < .001$ , while Frustration increased by 12%, *Student’s*  $t(334) = -8.51, p < .001$ .

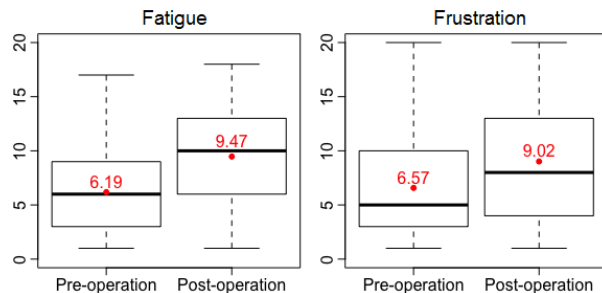


Figure 1: Boxplots with mean points of pre- and post-operation Fatigue and Frustration

**Cognitive Workload.** Operator cognitive workload was assessed across the six TLX factors (Figure 2). The calculated RTLX score was 44.55 (SD = 28.1).

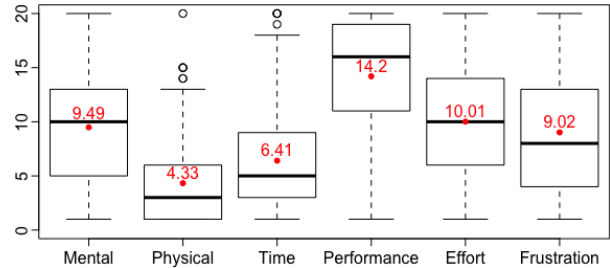


Figure 2: Boxplots with mean points of TLX cognitive workload factors

## 4.2 Instrument Validation

In the following section, we discuss our measures for establishing the reliability and generalizability of the COSS for studying cyber operations.

**Internal Validity.** For a research instrument to be considered reliable, it is expected that repeated studies of the same phenomena with the same instrument should produce the same or similar results. To show reliability of the COSS, we compare measurements between the four study sites. While these sites have individual differences, they are for the most part nearly identical in terms of operators, tools and techniques, and their missions. A one-way analysis of variance (ANOVA) shows no differences between sites for all stress factors (Table 2). Operators at all four study sites reported the same levels of fatigue, frustration, and cognitive workload.

Cognitive Stress Factors		N	Df	F	p
Fatigue	Pre-operation	357	3	0.53	.644
	Post-operation	356	3	0.46	.714
Frustration	Pre-operation	335	3	0.78	.503
	Post-operation	357	3	0.72	.541
Cognitive Workload	Mental demand	356	3	1.62	.184
	Physical demand	356	3	0.06	.982
	Time pressure	356	3	2.12	.097
	Performance	352	3	0.60	.615
	Effort	356	3	0.14	.936

Table 2: ANOVA comparisons of fatigue, frustration, and cognitive workload between study sites

An additional assessment for reliability is with the TLX itself. The TLX is a mature and widely used research instrument, and the six TLX factors (mental demand, physical demand, time pressure, performance, effort, frustration) tend to reliably correlate [10]. In our studies, five of the six TLX factors correlated in an expected way with the exception of performance (Table 3). The only significant relationship to self assessment of performance was to frustration. We believe the mission-critical nature of the environment we studied in which failure is not an option had an impact on this measure. This effect is further discussed in [17].

	Mental	Phys.	Time	Perf.	Effort	Frust.
Mental	–	.479*	.547*	-.034	.686*	.468*
Phys.	.479*	–	.541*	-.012	.486*	.334*
Time	.547*	.541*	–	-.022	.509*	.429*
Perf.	-.034	-.012	-.022	–	-.009	-.315*
Effort	.686*	.486*	.509*	-.009	–	.469*
Frust.	.468*	.334*	.429*	-.315*	.469*	–

Table 3: Pearson coefficient correlations ( $r$ ) between TLX factors (complete cases,  $N = 350$ ,\*  $p < .001$ )

**External Validity.** For an instrument to be considered generalizable, it is expected that it can be applied to other similar problems. While there are few previous studies of cyber operations that use the TLX, the RTLX scores from our four studies (RTLX = 44.55, SD = 28.1) align with the results of related previous work in cyber defense, such as Champion et al.’s study [2] of incident response using CyberCog (RTLX = 56.94, SD = 21.4) and Greenlee et al.’s report [6] on cyber defense tasks related to network triage (RTLX = 51.94, SD = 14.2) and incident escalation (RTLX = 40.04, SD = 11.6).

## 5 Lessons Learned and Conclusions

Our case study of NSA tactical cyber operators provided valuable feedback and validation about the design and use of the COSS. We share our experiences through a discussion of lessons learned.

**Participant Motivation.** At the start of our first study, participants seemed doubtful about the value of this research or its ability to motivate change. Initial responses to calls for participation were unenthusiastic, and it required an extraordinary effort to recruit. However, participation attitudes changed significantly after we reported our initial findings to the operator population. This report was also the basis of changes in operational policies that directly impacted many of our participants.

We believe presenting results back to the participants was very important for two reasons: First, it demonstrated that their participation mattered and helped generate actionable results. Second, by sharing research results back with the participation population, we demonstrated that their participation was valuable. Too often participants in research (especially local, corporate research) never see the results of the work. The presentation of early research results motivated other locations to participate. They saw that the research had results that mattered and affected changes in policies and procedures that mattered to them, and wanted to be sure that they were represented. Participation rates at these locations were very good; two locations even setup an informal ‘competition’ for number of participants.

**Other Design Considerations.** Environmental differences and specific workplace dynamics in cybersecurity

operations may have different effects on stress. This is why we recommend including additional contextual factors as needed. In an pilot version of the COSS, we asked participants to report the number of caffeinated drinks they consumed that day. We thought that caffeine intake might affect stress or fatigue reported during an operation. When we found no discernible correlation, we discarded the question from later versions of the survey in favor of other questions that were more relevant.

Another example of how the initial pilot study influenced the final design of the COSS is in measure of teaming. Originally, we had designed the survey as an individual measurement. Participant perception about the effects of teaming on frustration and performance led us to add a measure of team synergy for providing additional context. However, this singular measure falls short of a true team assessment.

The COSS survey questions are self-reported and these can be impacted by exogenous factors. We do not presume to explain why people report feeling the way they do, just how stressed they report feeling. In fact, if our measure is impacted by the workplace, it makes the COSS more relevant for organizational experiments because it can show how a change in the workplace affected (or failed to affect) employees and their work.

Another use for the COSS is as a method measuring and evaluating effects of stress interventions in the cybersecurity environment. One of our motivations for our data collection timeline was to capture the state of stress in current operations before major organizational changes took place. We now have a comprehensive baseline to which to compare in future work.

**Future Work.** A natural follow-on to this work is to extend to use objective data to detect or predict fatigue, frustration, and cognitive workload. While subjective data is relatively quick and easy to collect, it is also biased towards the perspective of the participant, often reflective in nature, and requires active participation. Objective measures, such as biometric data using heart rate or eye movement, can be measured passively, reducing the need to interrupt the observed task with potential impact on operations. Additionally, real-time analysis of this data could be incorporated into automated risk assessments that can measure changes in the operator and evaluate risk for that particular operation, rather than relying on generalized heuristic-based models of risk.

**Conclusion.** In this paper, we presented the Cyber Operations Stress Survey (COSS), an effective research instrument that can support the study and understanding of stress in tactical cyber operations. Cybersecurity researchers and practitioners alike can utilize the COSS to various ends. We look forward to seeing how researchers use the COSS in their own cybersecurity operations.

## References

- [1] D. Botta, R. Werlinger, A. Gagné, et al. 2007. Towards Understanding IT Security Professionals and Their Tools. *Proceedings of ACM Symposium on Usable Privacy and Security 2007*, 100-111.
- [2] M. Champion, P. Rajivan, N. J. Cooke, S. Jariwala. 2012. Team-based cyber defense analysis. *Proceedings of the IEEE International Multi-disciplinary Conference on Cognitive Methods in Situational Awareness and Decision Support*, 218-221.
- [3] W. Chappelle, K. McDonald, J. Christensen, et al. 2013. Sources of occupational stress and prevalence of burnout and clinical distress among U.S. Air Force cyber warfare operators. Technical Report no. AFRL-SA-WP-TR-2013-0006, Air Force Research Lab, Wright-Patterson AFB, OH, USA.
- [4] R. C. Dreifelbis, J. Martin, M. D. Covert, D. W. Dorsey. 2018. The Looming Cybersecurity Crisis and What it Means for the Practice of Industrial and Organizational Psychology, *Industrial and Organizational Psychology*, Vol. 11, No. 2, 346-365.
- [5] J. Dykstra & C. L. Paul. 2015. Stress and the Cyber Warrior: Workload in a Computer Operations Center. *Journal of Sensitive Cybersecurity Research and Engineering*, Vol. 3, No. 1, 1-23.
- [6] E. T. Greenlee, G. J. Funke, J. S. Warm, et al. 2016. Stress and workload profiles of network analysis: not all tasks are created equal, *Proceedings of the AHFE International Conference on Human Factors in Cybersecurity*, 153-166.
- [7] R. A. Grier. 2015. How high is high? A meta-analysis of NASA-TLX global workload scores. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 59, No. 1, 1727-1731.
- [8] J. Haney & C. L. Paul. 2016. Toward integrated tactical operations for red/blue network defense team. *Proceedings of the IEEE Conference on Military Communications Conference*.
- [9] S. G. Hart. 2006. NASA-task load index (NASA-TLX): 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 50, No. 9, 904-908.
- [10] S. G. Hart & L. E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, Vol. 52, 139-183.
- [11] J. M. Hektner, J. A. Schmidt, M. Csikszentmihalyi. 2007. *Experience Sampling Method: Measuring the Quality of Everyday Life*. Sage Publications, Thousand Oaks.
- [12] J. F. Hopstaken, D. van der Linden, A. B. Bakker, M. A. J. Kompier. 2014. A multifaceted investigation of the link between mental fatigue and task disengagement. *Psychophysiology*, Vol. 52, No. 3, 305-315.
- [13] K. M. Kowalski-Trakofler, C. Vaught. 2003. Judgment and decision making under stress: an overview for emergency managers. *International Journal of Emergency Management*, Vol. 1, No. 3, 278-289.
- [14] W. C. Li, F. C. Chiu, Y. S. Kuo, K. J. Wu. 2013. The Investigation of Visual Attention and Workload by Experts and Novices in the Cockpit. *Proceedings of International Conference on Engineering Psychology and Cognitive Ergonomics: Applications and Services*, 167-176.
- [15] C. L. Paul. 2014. Human-centered study of a network operations center: experience report and lessons learned. *Proceedings of the ACM CCS Workshop on Security Information Workers*, 39-42.
- [16] C. L. Paul & J. Dykstra. 2017. Fatigue, Frustration, and Cognitive Workload in Remote Operations across NSA Cryptologic Centers. *Journal of Intelligence Community Research and Development*, #450, September 2017.
- [17] C. L. Paul & J. Dykstra. 2017. Understanding Operator Fatigue, Frustration, and Cognitive Workload in Tactical Cybersecurity Operations. *Journal of Information Warfare*, Vol. 16, No. 2, 1-11.
- [18] M. Pomerleau. 2017. Exclusive peak inside Cyber Command's premiere annual exercise: Cyber Flag. *Army Times*, June 30 2017. <https://www.armytimes.com/home/2017/06/30/exclusive-peek-inside-cyber-commands-premiere-annual-exercise-cyber-flag/>
- [19] S. W. Samn & L. P. Perelli. 1982. Estimating aircrew fatigue: a technique with application to airlift operations, Technical Report no. SAM-TR-82-21, School of Aerospace Medicine, Brooks AFB, TX, USA.
- [20] B. Sawyer, V. Finomore, G. Funke, et al. 2014. Cyber vigilance: effects of signal probability and event rate. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 58, No. 1, 1771-1775.
- [21] D. van der Linden, G. P. J. Keijsers, P. Eling, R. van Schaijk. 2005. Work stress and attentional difficulties: An initial study on burnout and cognitive failures. *Work & Stress*, Vol. 19, No. 1, 23-36.
- [22] P. T. van Katwyk, S. Fox, P. E. Spector, E. K. Kelloway. 2000. Using the Job-related Affective Well-being Scale (JAWS) to investigate affective responses to work stressors, *Journal of Occupational Health Psychology*, Vol. 5, No. 2, 219-230.

## Appendix

See figures on the following pages.



