

Towards systematic and objective evaluation of police officer performance in stressful situations

J Bertilsson^{a,b}, DC Niehorster^{b,c,d,e}, PJ Fredriksson^a, M Dahl^d, S Granér^d, O Fredriksson^a, JM Mårtensson^a, M Magnusson^b, PA Fransson^b and M Nyström^c

^aPolice Region South, Swedish Police Authority, Malmö, Sweden; ^bDepartment of Clinical Sciences, Lund University, Lund, Sweden; ^cLund University Humanities Lab, Lund, Sweden; ^dDepartment of Psychology, Lund University, Lund, Sweden; ^eLund University Cognitive Science, Lund University, Lund, Sweden

ABSTRACT

To ensure a continuous high standard of police units, it is critical to recruit people who perform well in stressful situations. Today, this selection process includes performing a large series of tests, which still may not objectively reveal a person's capacity to handle a life-threatening situation when subjected to high levels of stress. To obtain more systematic and objective data, 12 police officers were exposed to six scenarios with varying levels of threat while their heart rate and pupil size were monitored. The scenarios were filmed and six expert evaluators assessed the performance of the police officers according to seven predefined criteria. Four of the scenarios included addressing a moderate threat level task and the scenarios were executed in a rapid sequence. Two further scenarios included a familiar firearm drill performed during high and low threat situations. The results showed that there was a large agreement between the experts in how they judged the performance of the police officers ($p < 0.001$). Performance increased significantly over tasks in four of the seven evaluation criteria ($p \leq 0.037$). There was also a significant effect of pupil size ($p = 0.004$), but not heart rate, when comparing the different sequential scenarios. Moreover, a high level of threat considerably impaired the motor performance of the police officers during the firearms drill ($p = 0.002$). Finally, the pupil seemed to systematically dilate more when a threat appeared immediately than with a delay in the scenarios ($p = 0.007$). We conclude that systematic and quantitative judgments from experts provide valuable and reliable information about the performance of participants in realistic and stressful policing scenarios. Furthermore, objective physiological measures of heart rate and pupil size may help to explain and understand why performance sometimes deteriorates.

ARTICLE HISTORY

Received 16 April 2019
Accepted 2 September 2019

KEYWORDS

Perception; motor control; tactical implementation; stress response; performance evaluation

Introduction

Police work is challenging and requires a plethora of skills to handle the wide range of situations that may occur during everyday work. According to studies cited in Anshel (2000), being a police officer is one of the most stressful jobs in the world (Anshel, 2000; Gyamfi, 2012). The stress may manifest itself as physical stress, which includes physical activity such as running, lifting and carrying, and psychological or psycho-social stress, which may be triggered by critical incidents such as threatening situations where the police officers feel they cannot control the situation

CONTACT J Bertilsson  Johan.Bertilsson@med.lu.se  Department of Clinical Sciences, Lund University, Lund S-221 85, Sweden

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

(Anderson, Litzenger, & Plecas, 2002). Stress can significantly impair the performance of police officers and, during long-term exposure, lead to conditions such as burnout and long-term illness (Gyamfi, 2012; Savic, 2015; Siddle, 1995; Yaribeygi, Panahi, Sahraei, Johnston, & Sahebkar, 2017). Police-specific simulated training exercises have been shown to induce a substantial amount of stress, and influence tasks relevant to police performance (Nieuwenhuys & Oudejans, 2011; Nieuwenhuys, Savelsbergh, & Oudejans, 2012).

On a physiological level, the stress response is a strong activation of the sympathetic nervous system (SNS), part of the autonomic nervous system (ANS), due to some kind of experienced threat. Emotions of fear or anxiety are connected to the stress response but likely stem from conscious working memory circuitry (LeDoux & Pine, 2016). The stress response has certain physiological effects on our central and peripheral nervous systems, affecting our neurological systems to various degrees depending primarily on the kind of threat and how severe it is perceived to be (LeDoux, 2000; Wang et al., 2018). Physiologically, stress can manifest itself as dilated pupils and bronchial tubes, increased heart rate, breathing, blood pressure, muscle strength and awareness, etc. (Ulrich-Lai & Herman, 2009). Stress has been shown to affect neurocognition (Dawes et al., 2014). The effects of repeated stress exposure have been less examined but reports suggest ramifications on cognition (Yuen et al., 2012) and that stress exposures in close succession may cause an increasing stress response level (J. Bertilsson et al., 2019).

Being able to handle stressful situations is important for any police officer but particularly for members of police special units, such as special weapons and tactics (SWAT) teams, who are exposed to potentially dangerous and threatening situations to a larger extent than the average police officer. However, also regular patrol officers face high risk of injury or death (Meyerhoff et al., 2004) because they typically arrive first to a situation, before the conditions on the site have been determined (Bertilsson, Petersson, Fredriksson, Magnusson, & Fransson, 2017; Greenberg, 2007; Petersson, Bertilsson, Fredriksson, Magnusson, & Fransson, 2017). An elite SWAT unit fills a critical role in minimizing the risks of injuries among police officers, the civilian population and the perpetrators alike while resolving anticipated or protracted high threat situations. Since these situations commonly are associated with high levels of stress, a crucial aspect when recruiting members to a police force is to assess how the recruits perform in situations where stressors are present (Bertilsson et al., 2013).

One route towards more systematic and objective assessment of police officer performance is through the use of structured expert-evaluations in combination with physiological markers. Evaluation by experts could include viewing video recordings of the recruits while they perform a series of tasks and assessing their performance according to a systematic protocol with questions relevant to the position the recruits are applying for. Such questions could be related to how they handle a situation in terms perception, communication, voice control, motor control, spatial and temporal tactical implementation. Physiological markers, on the other hand, may provide a means to objectively estimate the level of stress in response to a threatening situation (Anderson et al., 2002; Bertilsson et al., 2013; Carroll et al., 2000; Freyschuss, Hjemdahl, Juhlin-Dannfelt, & Linde, 1988; Sawai, Ohshige, Yamasue, Hayashi, & Tochikubo, 2007). In the context of police officer recruitment, this is useful both in terms of measuring the stress response of applicants in a test situation, but also through pre-testing ensuring that the test situations trigger a desired level of stress.

Due to the strong relationship between stress and heart rate, several studies have used heart rate monitoring to quantify stress in response to threatening situations (Atkins & Norris, 2004; Bertilsson et al., 2013; Vonk, 2008). In a study on police patrol officers, Anderson et al. found a systematic relationship between heart rate to both physical and psychological stress (Anderson et al., 2002). Examples of situations that led to an increase in heart rate were hands on a holstered gun and discussions with people suspected of crime. The relationship between stress and heart rate can also be found in more controlled environments such as laboratory studies, for instance using the Trier Social Stress Tests (TSST), where participants are exposed to public speaking and

arithmetic problem solving (Kirschbaum, Pirke, & Hellhammer, 1993). The TSST can lead to heart rate increases of about 25% in a population of University students (Kirschbaum et al., 1993).

Psychological stress has also been shown to cause dilation of the pupil, even though this relationship seems to be less studied than that between stress and heart rate (Beatty & Lucero-Wagoner, 2000). One example of the systematic relationship between pupil size and stress is a study on car driving, where pupil size could be used with a relatively high accuracy (79.2%) to distinguish low-stress situations from high-stress situations (Pedrotti et al., 2014).

Currently, the differences between the training methods and tests used for assessing the police officer's perceptive, cognitive and motor skills in connection with a stress response are very large, and the pros and cons of the different methods used are extensively debated. Some argue, for example, that the biological stress marker heart rate is not related to performance (Eamonn, Daugherty, & Arnetz, 2019) whereas others argue the opposite (Siddle, 1995). Thus, to more conclusively examine the effects of stress on performance, research is needed that details how to measure the level of stress and how to assess objectively the effects of stress during training and test settings that include different levels of threat (Bertilsson, Fredriksson, Piledahl, Magnusson, & Fransson, 2014). Hence, one objective of this study is to expand the current knowledge regarding this highly debated, but also very important, subject on how to handle the effects of high levels of stress, which in extreme situations can be a matter of life and death, and how to measure it with appropriate physiological biomarkers.

The issue of stress effects on performance is still somewhat taboo and has at times been attributed to poor personality traits and mental weakness (Paton & Violanti, 2008). However, since the stress response is an unconscious neurological reflex in all humans and most mammals (Nesse, Bhatnagar, & Ellis, 2016), this is not something that can be selected away by only using individuals supposedly lacking these traits. This means that after selection of the most experienced and resilient in tests, it is

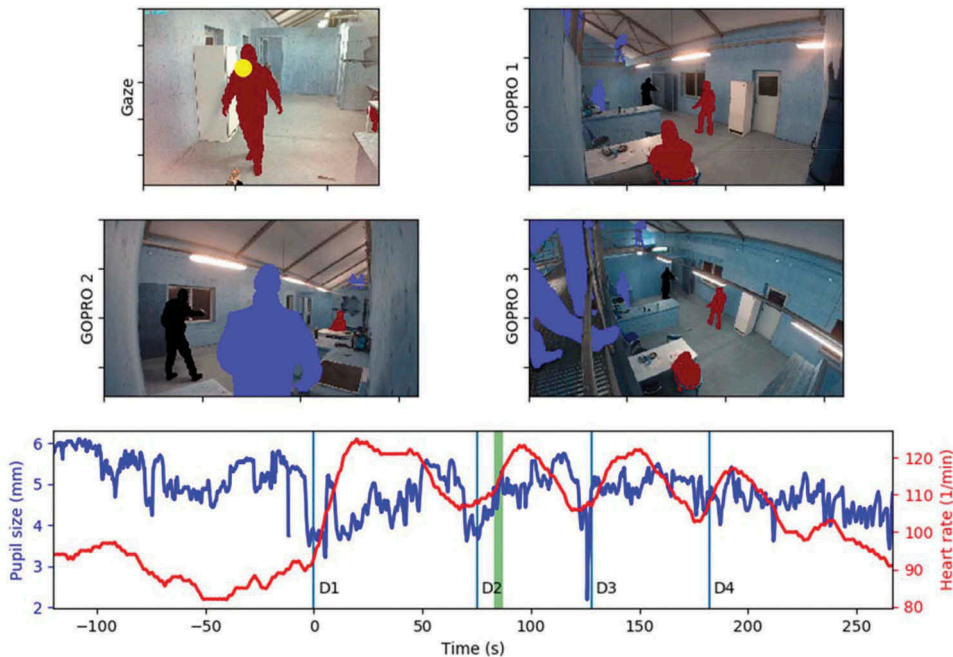


Figure 1. The participant's performance of the tasks was monitored by the SMI eye-tracking glasses (recording pupil diameter and a first-person view), three go-pro cameras monitoring the scene and a bioharness (recording heart rate). The study participant is marked in black, his gaze location is a yellow dot, the scenario figurants are marked in dark-red and the scenario instructors are marked in blue. The green bar in the heart rate/pupil size diagram shows for what period during the test sequence the images displayed originate. D1 – D4 in the recording stands for performing tasks 1 – task 4.

still necessary to ascertain that these individuals can deal with the consequences of the stress response, which may require optimized training of properly selected techniques, tactics and use of adapted equipment to diminish the negative effects (Bertilsson et al., 2017; Petersson et al., 2017). That said, sometimes training and experience might still not be enough to make someone able to handle high-stress situations in real life (Dahl, Granér, Fransson, Bertilsson, & Fredriksson, 2018). To enable us to know what to demand of applicants to the police force or special units, and how to improve selected abilities further, we need to learn more about our stress response and its effect in different settings. This study is a stepping stone in this work to improve human ability to respond in a risk-minimizing way for all involved in difficult or dangerous situations.

There are studies of the psychological correlates of behaviors under conditions of threat but this study focuses on the physiological markers. It is largely unknown how physiological measures of heart rate and pupil size relate to the performance of police officers over repeated tasks and at different levels of threat. The aim of this study was to investigate how exposure to repeated threats as well as different levels of threat influences the performance and physiological measures (heart rate and pupil size) of police officers in realistic policing scenarios. The performance was assessed by experts in terms of perceptual, cognitive, and motor skills during six test scenarios.

Methods

Experiments were performed in accordance with the Helsinki declaration and the recorded data were handled according to the procedures approved by the Scientific Ethical Committee at Lund University, Sweden (Dnr 2014–36). All participants provided written informed consent. Analyses focusing on a different aspect of the same dataset are presented in (Bertilsson et al., 2019).

Participants

Twelve healthy male subjects participated in the study ($M = 30.7$, $SD = 3.2$ years). However, due to recording malfunctions the heart rate recordings from one of the participants and pupil diameter recordings from another participant were excluded from analysis. The data were collected while

Table 1. Scenario description and duration of tasks.

Task	Scenario description	Duration (Mean (SD)) (seconds)
1.	Immediate moderate threat encounter scenario; Hostage situation where a person stands behind another person and holds a knife to their throat. The hostage-taker releases the hostage when the police officer enters the room, and walks towards the police officer.	26 (8)
2.	Delayed moderate threat encounter scenario; When the police officer enters the room, a person sitting calmly at a table is in plain sight. Suddenly a person with a knife appears from behind a fridge and moves toward the police officer while yelling angrily.	34 (22)
3.	Immediate moderate threat encounter scenario; Hostage situation where a person holds another person who sits at a table from behind and holds a gun to their head. When the police officer enters, the hostage taker turns the gun toward the officer.	25 (6)
4.	Delayed moderate threat encounter scenario; Compliant persons in a room with two dangerous objects visible in plain sight. When the police officer opens the door, one person is standing to his right and suddenly another curious person appears from behind the opened door.	136 (48)
5.	Immediate high threat encounter scenario; A man points a gun at the police officer from a corner of a room. The gun threat must be handled by shooting, but the police officer's gun has a jam that needs to be cleared before the police officer can use his pistol.	11 (4)
6.	No threat encounter scenario; Live fire jam-clearing drill test on a firing range.	≈120

the participants performed tasks in test scenarios. All participants were experienced police officers who had at least 5 years of prior experience with fieldwork. A prerequisite for participation was to be in good health. All participants had normal or corrected-to-normal vision (with contact lenses).

Equipment

Heart rate and body posture were recorded during tasks 1–6 with a Zephyr Bioharness® 2.0 (Zephyr technology cooperation, Annapolis, USA), and the pupil diameter was recorded during tasks 1–5 with eye tracking glasses from SensoMotoric Instruments (SMI), Berlin, Germany. Additionally, the participants' performance during tasks 1–6 was monitored by two or three Go-Pro™ 3.0 cameras, which also recorded sound. These were placed such that the actions of all actors in the scenario were always recorded by at least one camera, see [Figure 1](#). Before the task sequence started, a three-point calibration of the SMI glasses was performed followed by manual inspection of the accuracy by asking the participant to fixate another set of points on a wall. Additionally, both the SMI glasses and Bioharness system detected when the participants leaned forward to equip themselves with a belt and holster. This enabled time-synchronization of the heart rate recordings made with the Bioharness and the pupil diameter recordings provided by the SMI eye-tracking glasses. The Bioharness system sampled the ECG activity at 250 Hz and the SMI glasses recorded the pupil diameter at 30 Hz.

Procedure

Six different tasks were performed by the participants, including no threat, moderate threat, and high threat scenarios (see [Table 1](#) for a detailed description). One task was performed during the morning (task 6) and five tasks during the evening the same day (tasks 1–5), where tasks 1–4 were performed in a sequence with a brief rest between tasks of about 33 s ($M = 33.2$, $SD = 9.4$ s). Task 5 was performed about 4 h after the task 1–4 sequence.

The effects of repeatedly performing stressful tasks at moderate threat levels were evaluated by comparing questionnaire scores from six experts in police officer training (see 'Analysis' section for details), recorded heart rate and pupil diameter (see 'Equipment' section below), during tasks 1–4 that were performed in sequence with a brief rest between tasks. The effects of performing tasks under different levels of threat were determined by comparing expert review scores of motor performance when executing the identical firearm jam-clearing drill at no threat (task 6) and at high threat (task 5). Moreover, a more detailed comparison between performance under moderate (tasks 1–4) and high (task 5) threat was done by comparing performance as measured by expert review scores, and by comparing evoked stress reactions as by measuring heart rate and pupil diameter. Additionally, it was also determined if any of the parameters used to assess performance were affected by whether a threat appeared immediately at the onset of the task or if it appeared after a delay. Based on recorded heart rate, tasks 1–6 were in retrospect categorized to produce threat at low (mean 95.9 bpm (SD 9.9)), moderate (mean 123.7 bpm (SD 10.8)) and high (mean 141.7 bpm (SD 30.0)) levels. The policing actions performed by the participant during all scenarios involved at most slow walking (i.e., no physical strain likely to cause increased heart rate).

Before starting each task (task 5 and 6) or task sequence (task 1–4), the participants put on the recording equipment, which was calibrated and synchronized. The participants were also equipped with a SigSauer adapted for Simunition cartridges and a pepper spray canister that for training purposes contained only water. The morning task (task 6) was performed outdoors in daylight. All of the five evening tasks (task 1–5) were preceded by the participant standing for about 30 s in a dimly lit (4.6 Lux) anteroom. All evening tasks started when the participants opened a door between the anteroom and a scenario room, where a specific scenario started to

play out immediately. When the door to the more brightly lit scenario room was opened, the participant was exposed to an illumination of 191 Lux. When completely inside the scenario room, the illumination increased to about 385 Lux. When a scenario instructor judged the policing task to be completed, he terminated the task with a verbal ‘abort’ command and, if performing the task sequence, the participant returned to the anteroom for a brief rest before the next task commenced. The participant received no detailed instructions about the scenario to address and was merely asked to deal with the situation.

Analysis

The performance of each participant during each of the tasks was rated by six experts in police officer training practices. The experts reviewed the performance using the video recorded with the front-facing camera on the SMI eye-tracking glasses, which was shown to the raters with the participants’ gaze overlaid on the video. The experts also reviewed the films from the three Go-Pro cameras displaying an overview of the scene from different viewpoints, see Figure 1. However, the participant’s heart rate and pupil diameter values were not revealed to the experts. The experts were allowed to watch the participant’s performance as many times as they wanted. The order in which the experts rated videos from the different participants and tasks was randomized. The experts scored the participants’ performance using a custom-made questionnaire (Appendix A), targeting seven different aspects of performance. These categories were; I – perception, II – verbal content, IIIa – verbal voice control, IIIb – general motor control, IIII -spatial and V – temporal tactical implementation, and VI – overall situational control. After determining that the experts’ responses were highly correlated with each other, the average scores from all six experts were calculated and used in the statistical analyses.

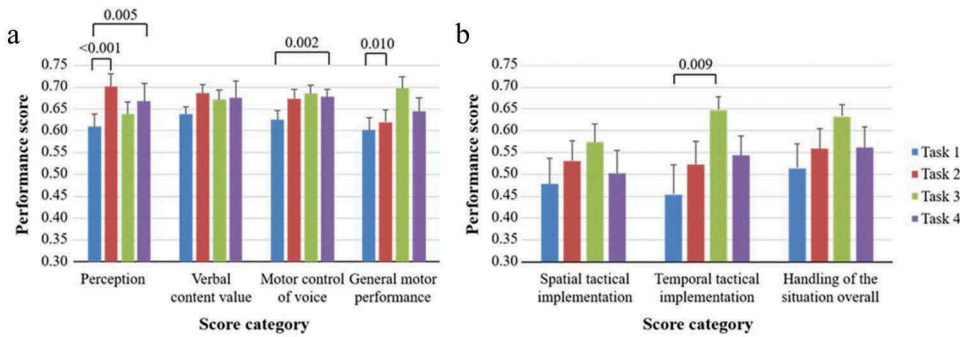


Figure 2. (a) Perception, verbal and motor control performance scores (I-IIIb). (b). Spatial and temporal tactical implementation and overall situation control (IIIIV). The perception and motor control performance increased more clearly throughout all repeated tasks whereas mostly the spatial tactical implementation and overall situation control suffered a decline when encountering a more complex scenario in task 4.

Table 2. Effects of repetition on the expert review scores.

Performance category	Repetition
I. Perception	0.014 [8.5]
II. Verbal content	0.192 [1.9]
IIIa. Motor control of voice	0.004 [13.1]
IIIb. General motor control performance	0.037 [5.6]
IIII. Spatial tactical implementation	0.104 [3.1]
V. Temporal tactical implementation	0.023 [6.9]
VI. Handling of the situation overall	0.120 [2.8]

^aRepeated measures GLM ANOVA analysis of how the expert review scores for questionnaire categories I-VI were affected by main factor ‘Repetition’ during the repeated tasks.

The average heart rate and pupil diameter during each of the tasks were analyzed offline by a custom-made program. The **Onset** time of each task was defined as when the door between the anteroom and scenario room was opened, and thus, when the scenario was revealed to the participant. The **Offset** time of the task was defined as the moment when the scenario instructor terminated the task with a verbal ‘abort’ command.

Statistical analysis

The seven questionnaire categories (I–VI), the average heart rate, and the average pupil diameter during the four repeated moderate threat tasks 1–4 were analyzed using repeated measures GLM ANOVA. The main factor was ‘Repetition’ (Tasks 1, 2, 3, 4; d.f. 3). The repeated measures GLM ANOVA was used after ensuring that all dataset combinations analyzed in the study with this method produced residuals that had normal or close to normal distribution, thus validating the appropriateness of using a GLM ANOVA (Altman, 1991). Wilcoxon matched-pairs signed-rank tests (Exact sig. 2-tailed) were used for within-group post hoc comparisons, i.e., analyzing the differences between tasks.

A Spearman correlation analysis was performed to determine with what consistency the six experts in police officer training scored the performance of each of the test subjects and during each of the tests. Additionally, a reliability analysis was performed using the Cronbach’s Alpha test when evaluating the consistency in expert scoring. Moreover, Spearman correlation analysis was performed to determine any relationships between the expert’s performance scores and the biomarkers heart rate and pupil diameter during moderate threat (task 1–4) and during high threat (task 5).

In the post hoc analyses, p-values smaller than 0.0500, 0.0250 or 0.0167 were considered significant depending on the number of comparisons made, in line with Bonferroni correction procedures. The Shapiro–Wilk test revealed that some datasets were not normally distributed and that a normal distribution could not be obtained by log-transformation. Thus, non-parametric statistical methods able to

Table 3. Effects of repetition on the heart rate and pupil diameter.

Physical manifestations	Repetition
Heart rate	0.217 [1.7]
Pupil diameter	0.004 [14.1]

^aRepeated measures GLM ANOVA analysis of how the mean heart rate and pupil diameter were affected by main factor ‘Repetition’ during the repeated tasks.

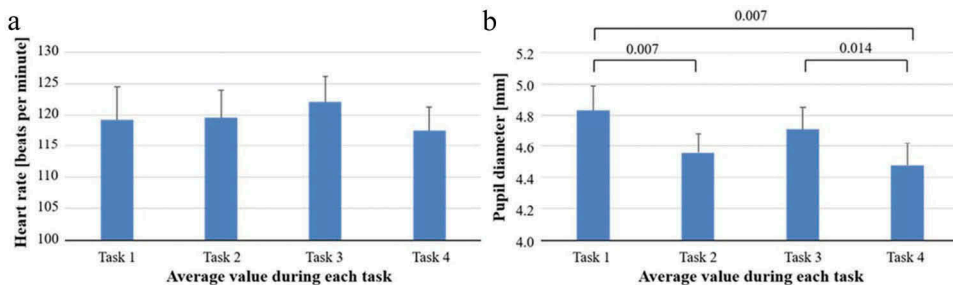


Figure 3. (a) Average heart rate during tasks 1–4 (moderate threat tasks). No significant difference in heart rate was found between repeated moderate threat tasks. (b) Average pupil diameter during tasks 1–4 (moderate threat tasks). The pupil diameter was significantly larger during tasks where the threat appeared immediately (tasks 1 and 3) compared with tasks with delayed threats (tasks 2 and 4).

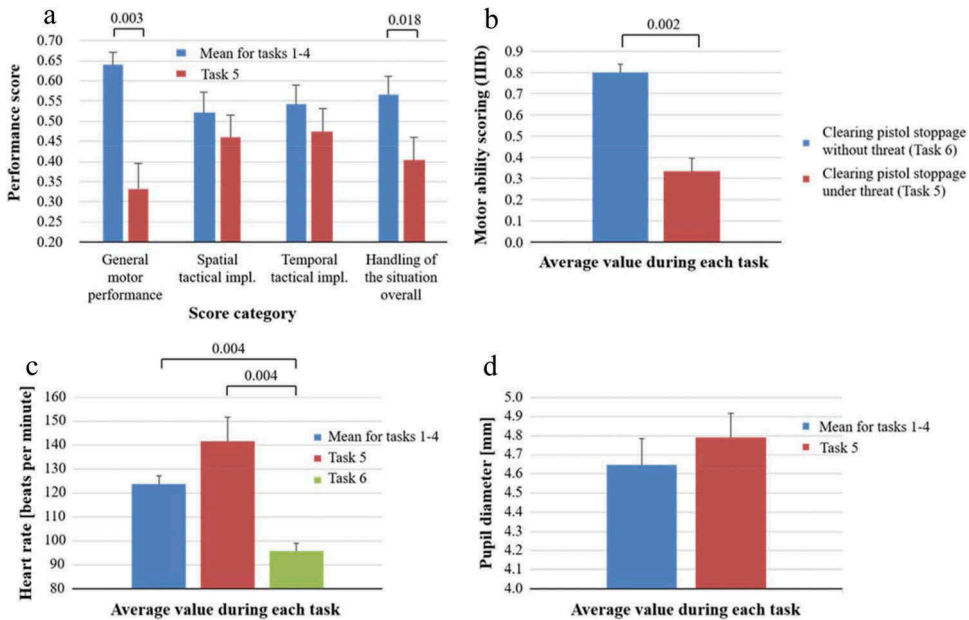


Figure 4. (a) Having to handle a task with higher threat caused a significant decrease in motor control performance (IIIb) and in overall situation control (VI). (b). The general motor control performance (IIIb) was significantly poorer when resolving/clearing a pistol jam during high threat (task 5) compared with no threat (task 6). (c). The average heart rate was significantly higher when submitted to moderate threat (Tasks 1–4) and a high threat (task 5) compared with performing a task without threat (task 6). Despite large difference in mean heart rate between moderate and high threat tasks, due to large individual differences during the high threat task this effect did not reach significance. (d). The pupil diameter was about the same under high threat (Task 5) as under moderate threat (tasks 1–4).

handle comparisons of individual datasets with non-normal distributions were used in all post hoc statistical evaluations (Altman, 1991).

Results

Properties of expert evaluations

The Spearman correlation analysis between the six individual expert evaluations, as manifested by the scoring form, revealed a strong consistency between the ratings of the individual experts ($p < 0.001$, $r \geq 0.468$). The reliability analysis revealed a Cronbach's Alpha value of 0.890.

Effects of task repetition on the expert review scores

When repeatedly performing tasks that induced moderate threat in a sequence, performance significantly increased over tasks for the perception (category I, $p = 0.014$), motor control of voice (IIIa, $p = 0.004$), general motor control (IIIb, $p = 0.037$), and temporal tactical implementation (V, $p = 0.023$) categories (Figure 2, Table 2).

The post hoc evaluation suggested that the test subject's performance systematically improved by repetition in verbal motor control (category IIIa, $p = 0.002$ – between tasks 1 vs. task 4), in general motor control (category IIIb, $p = 0.010$ – between tasks 1 vs. task 2), and in temporal tactical implementation (category V, $p = 0.009$ – between tasks 1 vs. task 3). However, statistical findings for perception (category I) suggest that tasks where the threat appeared immediately (tasks 1 and 3) produced a poorer perception than tasks with a delayed threat (tasks 2 and 4), ($p < 0.001$, task 1 vs. task 2; $p = 0.005$, task 1 vs. task 4)

Effects of task repetition on heart rate and pupil diameter

Repeatedly performing tasks under moderate threat had no significant effect on the average heart rate during tasks 1–4 (Table 3, Figure 3). However, the average pupil diameter during the tasks decreased significantly by repetition ($p = 0.004$).

A post hoc evaluation revealed no Bonferroni-corrected significant differences in heart rate between the moderate threat tasks 1–4 (Figure 3). However, a pattern emerged in the pupil diameter data, suggesting that the pupil diameter was significantly larger if the threat appeared immediately (tasks 1 and 3) compared with if the threat appeared with a delay (tasks 2 and 4) ($p < 0.007$, task 1 vs. task 2; $p = 0.007$, task 1 vs. task 4; $p = 0.014$, task 3 vs. task 4)

Effects of threat levels on the expert review scores

The high threat scenario (task 5) only allowed the experts to score four of the questionnaire categories, IIIb, III, V and VI (Figure 4(a)). Having to handle a high threat task caused a significant decrease in general motor control performance (IIIb, $p = 0.003$) and in overall situation control (VI, $p = 0.018$) compared with the average performance during moderate threat tasks (tasks 1–4).

Finally, when comparing expert review scores for performing a firearm jam-clearing drill during high threat (task 5) with performing the same task during no threat (task 6) (Figure 4 (b)), it was found that the test subject's motor control performance was rated significantly poorer in the high threat situation (IIIb, $p = 0.002$).

Effects of threat levels on heart rate and pupil diameter

The statistical evaluation revealed no significant differences between the mean heart rate values during the moderate threat tasks and during the high threat task 5 (Figure 4(c)). Moreover, the average size of the pupil (Figure 4(d)) during the moderate threat tasks 1–4 was not significantly different compared with the pupil diameter recorded during the high threat task (task 5).

Relationship between expert scores during moderate and high threat and recorded heart rate and pupil diameter

The Spearman correlation analysis revealed no significant relationships or trends between any of the performance scoring categories and the heart rate during the tasks with moderate threat (task 1–4) ($p \geq 0.125$, $r \leq -0.5$). However, during the high threat task (Task 5), trends suggest that higher heart rate levels were related to poorer general motor performance ($p = 0.082$, $r = -0.6$), poorer temporal tactical implementation ($p = 0.082$, $r = -0.6$) and a poorer handling of the situation overall ($p = 0.074$, $r = -0.6$)

The Spearman correlation analysis revealed one trend between a performance score and the pupil diameter during the tasks with moderate threat (task 1–4). During task 3, a larger pupil diameter tended to be related to better spatial tactical implementation ($p = 0.077$, $r = -0.6$). However, during the high threat task (Task 5), no significant relationship or trends between any of the performance scoring categories and the pupil diameter could be found ($p \geq 0.533$, $r \leq -0.2$).

Discussion

The requirements for police officers working in field duty are extensive. For officers working in the elite SWAT police units, the specific requirements are to perform well during physically strenuous and high threat situations. Today, selection of SWAT-recruits includes performing a series of tests, which still may not objectively reveal a person's true capacity to handle a life-threatening situation while submitted to high levels of stress. Hence, we need efficient tools to perform systematic and objective evaluations of police officer performance in stressful situations. The objective of the present study was to evaluate some of the factors that may affect performance

while executing stressful tasks. Another objective was to determine the reliability of using experts in police officer training to evaluate the performance of police officers during tryouts using a custom-made questionnaire, and to investigate whether the biomarkers heart rate and pupil diameter can explain some of the variation in performance.

Properties of expert evaluations

The statistical correlation analysis of the experts scoring of performance revealed a high consistency between the rating values obtained from different experts. The extensive experience of the experts is likely one of the key reasons that this, essentially subjective, method is still viable for practical use. Another key factor was likely that the experts were provided with the relevant information needed to conduct a proper review from the cameras and were allowed to repeatedly watch the films while judging performance. A novel and appreciated tool by the experts were the SMI eye-tracking glasses that displayed the participants' gaze overlaid on the video, and thus revealed continuously where the participants looked while performing the tasks. A complex problem is to determine what aspects of a performance are relevant to score. Our questionnaire focused on seven categories, but there is room for expanding it with more comprehensive and detailed questions, for instance, 'did the participants detect all harmful objects in the scenario room?'. The most obvious use of the gaze for the experts was to compare to what degree the participants tried to perceive possible threats by scanning the risk areas, followed by how soon they reacted to an already visible or suddenly presented threat.

A noteworthy limitation with using correlations to determine consistency of experts' evaluations is that correlation values only show that the experts consistently rate the same kind of participant performance as either good or bad. However, a correlation value does not take into account properties like offset and amplitude differences in the scoring of performance.

Effects of task repetition

Repeatedly performing tasks seemed to have potential to improve some of aspects of performance more than others. In the expert scores, the predominant aspects that improved with repetition were perception, motor control of voice, general motor control, and temporal tactical implementation (Table 2). Moreover, pupil diameter during the tasks decreased significantly by repetition. Hence, when exposed to a sequence of scenarios with different objectives and required actions, see Table 1, there is still potential to make improvements and adaptations. These improvements may include enhancing basic neurobiological skills like perception and motor control. However, it should be noted that the tasks that were repeatedly performed caused only a moderate heart rate increase (i.e., induced a moderate threat), to on average 120 bpm. Hence, it remains unclear whether skills like motor control also improve when repeatedly exposed to higher stress levels and whether stress adaptation at this moderate threat level transfers to high threat situations.

Effects of threat levels

Having to handle a high threat task caused a significant decrease in general motor control performance and in overall situation control compared with the average performance during moderate threat tasks (tasks 1–4). The most marked effect was found when comparing expert review scores for performing a firearm jam-clearing drill during high threat (task 5) with performing the same task during no threat (task 6). The motor control performance dropped significantly in the high threat situation. Hence, the study revealed that the level of stress may affect also an extensively trained complex motor skill (task 5. see Table 1) when under the influence of a strong stress response. This finding concurs with reports describing difficulties

using fine or complex motor skills when under a strong stress response (Applegate, 1976; Fairbairn & Sykes, 1942; Siddle, 1995).

The biomarkers heart rate and pupil diameter were not significantly different between moderate and high-stress levels, but the differences in average heart rate levels were large (see Figure 4 (c)). However, individual variations in heart rate were large during the high threat task, partly because at least one of the participants was sleepy, which may have caused him to show few signs of being affected by stress when performing the high-stress task. Moreover, the extensive field duty experience of the participants may have caused that the stress levels induced by the simulated high threat scenario was probably too low (induced an average of 142 bpm) to reveal the more dramatic effects of a high threat.

Effects of whether a threat appeared immediately or with a delay in the task scenarios

An intriguing finding is that the pupil diameter was systematically larger during scenarios where the threat appeared already at the start of the task compared with scenarios where the threat appeared with a delay. Moreover, the experts also scored participants significantly lower on perception skills when they had to address an immediate threat. Hence, the pupil diameter may reveal how participants respond to details in the scenarios they were exposed to.

The study's findings are of interest for several reasons. Firstly, the findings add new knowledge about how police officer's stress changes during various scenario conditions, which may have practical implications when developing police training and planning at strategic as well as operational/tactical levels. The knowledge can also be useful in after-action reviews of high-stress police interventions. Secondly, the study revealed that recordings of pupil activity could be used separately and in combination with heart rate as a reliable and recordable physiological stress marker, this also for research where the temporary effects of the stress response on performance are of interest. Finally, the study suggests that experts, when allowed sufficient time and sufficient access to footage of the testing situation, can assess police officer's performance also in details such as perception, cognition in verbal content, tactical implementation and motor control, to the level where the scorings reveal deficits in performance related to increased stress levels.

When considering aspects of practical usefulness, there exist today on the general market numerous comparatively low cost and easy to use equipment for measuring heart rate, which makes this stress recording approach viable also for local police organizations and for individual instructors who want to examine stress levels during and after training or tests. For example, heart rates due to the stress response around and over 145 bpm may explain why police officers evaluated display difficulties to perform fine and complex motor skills during tasks performed (Siddle, 1995). Another viable tool for evaluating the effects of stress responses was found to be expert evaluation of performance. However, the good performance of the expert evaluators was likely substantially enhanced by the fact that the test subjects in this study were continuously monitored during the tasks by three external cameras and one-eye tracking camera displaying where the test subject looked. Specifically, the films of participants view and gaze location may have been very useful when evaluating the perceptive abilities of the participants. If the pupil size changes are also made available to the instructors, then they may be better able to gauge how participants perceived the scenario subconsciously. However, the eye-tracking camera equipment used in this study presently costs upwards from several thousand US\$ at minimum, and demands assistance and technical expertise to handle. Thus, the device may only be suitable for use at larger regional police education and training centers.

Limitations

The small number of subjects ($n = 12$) limited the reliability of the statistical analysis of the relationship between the expert scores, the heart rate, and the pupil diameter and limited the opportunities to

investigate the effects of randomizing the order of tasks performed. However, an unexpected positive effect of the fixed test order turned out to be that differences in the responses of the biomarkers heart rate and pupil size, depending on scenario characteristics, were revealed. Moreover, human figurants were used in the tasks performed, which might have introduced differences in how each scenario played out for each of the test subjects, e.g., in verbal content and in movements made in the scenario room. However, this limitation was regarded of lesser relevance than the benefit that human figurants' can interact with the test subjects, which lends the scenario higher realism. That said, further research should be made on a larger number of subjects, using a randomized tests order, to determine a systematic role of design settings on the biomarkers responses to different situational characteristics.

Another limitation of this study is that all included subjects could be regarded as elite police officers. Thus, more studies should be made to determine baseline values in the performance rating forms, using larger groups of subjects consisting of a mixture of experienced and novice police officers before and after having received education about recommended approaches.

Finally, the study design gave restricted opportunity to determine the role of the individual psychological factors and their interactions, likely to have induced the stress in the test subjects. The vehicles commonly relied on to produce stress is a mix of fear of pain (Simunition paint marking cartridges), uncertainty (creates anxiety), surprises (startle response), own expectations and perceived own performance and the additional social stress factor due to the surveillance (cameras and measuring devices) and to being watched live while performing. However, we have in a previous study noted that the measuring equipment (heart rate) and performing the physical task while instructors watched gave a much lower mean heartrate than when performing the same task with fear of pain (Simunition), figurants and with reviewing witnesses (other instructors) (Bertilsson et al., 2013).

Conclusion

Systematic and quantitative judgements from experts provide valuable and reliable information about the performance of participants in realistic and stressful policing scenarios. The concept of guiding the experts' rating of the test subject's performance using a predesigned form that was sectioned into a set number of performance aspect categories, regarded of relevance in policing situations, turned out to be a successful approach, e.g., several of the scoring categories used revealed well when a test subject was affected by stress. Thus, this kind of approach can be recommended when evaluating recruits for police officer education, SWAT team membership, or for obtaining feedback verifying that a training method has produced the intended effects.

Since even a highly trained complex motor skill deteriorated significantly corresponding to a mean heart rate of 142 bpm during a high-threat (high-stress) scenario compared to repeated correct performance of the same motor skill corresponding to a mean heart rate of 95 bpm during no threat, the psychomotor effects of stress should become a more serious research subject in connection to police practices, tactics, training and equipment design in the future.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Medicinska Fakulteten, Lunds Universitet.

Notes on contributors

J Bertilsson, PhD and police sergeant, firearms and tactics instructor and worked as Chief use-of-force and self-defense instructor 2005 until 2015 at Skåne County Police Department, Sweden. Presently he work at Competence

Center South & East, The Swedish Police Authority, and received a PhD in Medical Science from Lund University, Sweden, in 2019. His research interests include perceptive, cognitive and motor skill performance depending on the effects of internal and external pressure like pre-training, psychological and physical stress.

DC Niehorster received a PhD from The University of Hong Kong, Hong Kong, in 2014. He is presently a researcher at the Lund University Humanities Lab. His research interests including eye tracking and eye tracking methodology, and visual perception.

PJ Fredriksson is a Sergeant and has professionally previously worked both as a firefighter and paramedic before he started working as a police officer. He held a position as a self-defense instructor at Scania County Police Department and as a developer of strategies concerning tactics, fire-arms, self-defense and training in the Scania County Police Department. Presently, he is working as a teacher in weapons and tactics at the Institute of Police Work at Malmo University, Malmo, Sweden. His research interests include performance when under pressure and ways to conceptualize the adaptation needed for different, common or dangerous situations.

M Dahl received a PhD in 2002 in psychology and holds a position as senior researcher at The Department of Psychology, Lund University. His research interest includes eyewitness memory, realism in confidence judgements and performance and decision-making under acute stress.

S Granér received his PhD from the Department of psychology at Lund University in 2010 and are currently a lecturer at the same department. His main areas of interest is cognitive and motor performance in high stress situations, in sport or other occupations where acute performance situations occur.

O Fredriksson, police sergeant, former team leader and field commander at South Sweden tactical unit and responded to several hundred high-risk field operations for 19 years. Also served in different staff positions in many emergency operations from 2013-2018. Presently he work at Swedish National Operations department (NOA) developing command and control systems for different levels in the command chain in emergency operations.

JM Mårtensson, acting chief inspector, for the forensic laboratory Police Region South and before that worked as staff member at operational management for Police Region South, Sweden. Previously a firearms, use-of-force instructor and operator at South Sweden tactical unit. Presently also active in the Swedish armed forces reserve as an intelligence officer.

M Magnusson received an MD in 1981 and a PhD in 1986 from Lund University, Sweden, became Associate Professor in Otorhinolaryngology in 1988 and a received a full Professorship in 1999. He presently holds a position as senior consultant and head of the division of Otolaryngology and is head of the section of Senses, Neuroscience and Psychiatry of the Department of Clinical Sciences, Lund University. His research interests involve inner ear and vestibular disorders, postural control and orientation.

PA Fransson received a PhD in Medical Science from Lund University, Sweden, in 2005 and the degree Associate Professor in 2009. He presently holds a position as senior researcher at the Department of Clinical Sciences, Lund University. His research interests include the human CNS, the sensory and motor systems and the functional decline or adaptation of these systems as an effect of physical and psychological stress, drugs and new training paradigms.

M Nyström received a PhD in Information Theory from Lund University, Sweden, in 2008, and become an Associate Professor in 2015. He currently works in the Lund University Humanities Lab. His research interests include vision, eye movements, and eye tracking.

ORCID

DC Niehorster  <http://orcid.org/0000-0002-4672-8756>

References

- Altman, D. (1991). *Practical statistics for medical research*. New York: NY: Chapman & Hall.
- Anderson, G. S., Litzenberger, R., & Plecas, D. (2002). Physical evidence of police officer stress. *Policing: An International Journal of Police Strategies & Management*, 25(2), 399–420.
- Anshel, M. H. (2000). A conceptual model and implications for coping with stressful events in police work. *Criminal Justice and Behavior*, 27(3), 375–400.
- Applegate, R. (1976). *Kill or get killed; Riot control, techniques, manhandling, and close combat for police and the military*. Boulder, Colorado, USA: Paladin Press, Paladin Enterprises, Inc.

- Atkins, V. J., & Norris, W. A. (2004). *Survival scores research project: FLETC*. Glynco, U.S: Department of Homeland Security.
- Beatty, J., & Lucero-Wagoner, B. (2000). *The pupillary system* (2nd ed. ed.). Cambridge, UK: Cambridge University Press.
- Bertilsson, J., Fredriksson, P. J., Piledahl, L. F., Magnusson, M., & Fransson, P. A. (2014). Opportunities and challenges of research collaboration between police authorities and university organizations. In F. Lemieux, G. Heyer, & D. K. Das (Eds.), *Economic development, crime, and policing; global perspectives* (pp. 163–180). New York, USA: CRC Press, Taylor & Francis Group. International Police Executives Symposium Co-Publication.
- Bertilsson, J., Niehorster, D. C., Fredriksson, P. J., Dahl, M., Granér, S., Fredriksson, O., ... Nyström, M. (2019). Stress levels escalate when repeatedly performing tasks involving threats. *Frontiers in Psychology, 10*(1562). doi:10.3389/fpsyg.2019.01562
- Bertilsson, J., Patel, M., Fredriksson, P. J., Piledahl, L. F., Magnusson, M., & Fransson, P. A. (2013). *Efficiency of simulated realistic scenarios to provide high psychological stress training for police officers*. London, UK: CRC Press, Taylor & Francis Group.
- Bertilsson, J., Petersson, U., Fredriksson, P. J., Magnusson, M., & Fransson, P. A. (2017). Use of pepper spray in policing: Retrospective study of situational characteristics and implications for violent situations. *Police Practice and Research, 18*(4), 391–406.
- Carroll, D., Harrison, L. K., Johnston, D. W., Ford, G., Hunt, K., Der, G., & West, P. (2000). Cardiovascular reactions to psychological stress: The influence of demographic variables. *Journal of Epidemiology and Community Health, 54*(11), 876–877.
- Dahl, M., Granér, S., Fransson, P. A., Bertilsson, J., & Fredriksson, P. (2018). Analysis of eyewitness testimony in a police shooting with fatal outcome – Manifestations of spatial and temporal distortions. *Cogent Psychology, 5*, 1487271.
- Dawes, D. M., Ho, J. D., Vincent, A. S., Nystrom, P. C., Moore, J. C., Steinberg, L. W., ... Miner, J. R. (2014). The neurocognitive effects of simulated use-of-force scenarios. *Forensic Science, Medicine, and Pathology, 10*(1), 9–17.
- Eamonn, A., Daugherty, A. M., & Arnetz, B. (2019). Differential effects of physiological arousal following acute stress on police officer performance in a simulated critical incident. *Frontiers in Psychology, 10*, 759. doi:10.3389/fpsyg.2019.00759
- Fairbairn, W. E., & Sykes, E. A. (1942). *Shooting to live*. London England: Oliver and Boyd.
- Freyschuss, U., Hjemdahl, P., Juhlin-Dannfelt, A., & Linde, B. (1988). Cardiovascular and sympathoadrenal responses to mental stress: Influence of beta-blockade. *The American Journal of Physiology, 255*(6 Pt 2), H1443–1451.
- Greenberg, S. F. (2007). Active shooters on college campuses: Conflicting advice, roles of the individual and first responder, and the need to maintain perspective. *Disaster Medicine and Public Health Preparedness, 1*(1 Suppl), S57–61.
- Gyamfi, G. (2012). Evaluating the effect of stress management on performance of personnel of Ghana police service *International Police Executive Symposium/United Nations Conference at: NYC*. United Nations.
- Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The ‘Trier social stress test’—A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology, 28*(1–2), 76–81.
- LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience, 23*, 155–184.
- LeDoux, J. E., & Pine, D. S. (2016). Using neuroscience to help understand fear and anxiety: A two-system framework. *The American Journal of Psychiatry, 173*(11), 1083–1093.
- Meyerhoff, J. L., Norris, W., Saviolakis, G. A., Wollert, T., Burge, B., Atkins, V., & Spielberger, C. (2004). Evaluating performance of law enforcement personnel during a stressful training scenario. *Annals of the New York Academy of Sciences, 1032*, 250–253.
- Nesse, R. M., Bhatnagar, S., & Ellis, B. (2016). Evolutionary origins and functions of the stress response system. In G. Fink (Eds.), *Stress: Concepts, cognition, emotion, and behavior* (Chap. 11, pp. 95–101). London, UK: Elsevier.
- Nieuwenhuys, A., & Oudejans, R. R. (2011). Training with anxiety: Short- and long-term effects on police officers’ shooting behavior under pressure. *Cognitive Processing, 12*(3), 277–288.
- Nieuwenhuys, A., Savelsbergh, G. J., & Oudejans, R. R. (2012). Shoot or don’t shoot? Why police officers are more inclined to shoot when they are anxious. *Emotion, 12*(4), 827–833.
- Paton, D., & Violanti, J. M. (2008). Law enforcement response to terrorism: The role of the resilient police organization. *International Journal of Emergency Mental Health, 10*(2), 125–135.
- Pedrotti, M., Mirzaei, M. A., Tedesco, A., Chardonnet, J. R., Mérienne, F., Benedetto, S., & Baccino, T. (2014). Automatic stress classification with pupil diameter analysis. *International Journal of Human-computer Interaction, 30*(3), 220–236.
- Petersson, U., Bertilsson, J., Fredriksson, P., Magnusson, M., & Fransson, P. A. (2017). Police officer involved shootings – Retrospective study of situational characteristics. *Police Practice and Research, 18*(3), 306–321.
- Savic, I. (2015). Structural changes of the brain in relation to occupational stress. *Cerebral Cortex, 25*(6), 1554–1564.

- Sawai, A., Ohshige, K., Yamasue, K., Hayashi, T., & Tochikubo, O. (2007). Influence of mental stress on cardiovascular function as evaluated by changes in energy expenditure. *Hypertension Research : Official Journal of the Japanese Society of Hypertension*, 30(11), 1019–1027.
- Siddle, B. K. (1995). *Sharpening the warrior's edge: The psychology & science of training*. Millstadt, IL: PPCT Research Publications.
- Ulrich-Lai, Y. M., & Herman, J. P. (2009). Neural regulation of endocrine and autonomic stress responses. *Nature Reviews. Neuroscience*, 10(6), 397–409.
- Vonk, K. D. (2008). *Police performance under stress*. Deerfield, IL: Hendon publishing company.
- Wang, Y., Zekveld, A. A., Wendt, D., Lunner, T., Naylor, G., & Kramer, S. E. (2018). Pupil light reflex evoked by light-emitting diode and computer screen: Methodology and association with need for recovery in daily life. *PLoS One*, 13(6), e0197739.
- Yaribeygi, H., Panahi, Y., Sahraei, H., Johnston, T. P., & Sahebkar, A. (2017). The impact of stress on body function: A review. *EXCLI Journal*, 16, 1057–1072.
- Yuen, E. Y., Wei, J., Liu, W., Zhong, P., Li, X., & Yan, Z. (2012). Repeated stress causes cognitive impairment by suppressing glutamate receptor expression and function in prefrontal cortex. *Neuron*, 73(5), 962–977.

Appendix A

Performance rating form

The performance of each participant during each of the tasks was rated by six experts in police officer training practices. The experts scored the participants' performance using a custom-made questionnaire, targeting seven different aspects of the performance.

I: Perception

How are the risk areas in the scenario scanned visually?

Assess with a cross between 0 = *substandard* & 1 = *excellent*

0 I ----- I 1

II: Verbal content value

Communication through content in what is said and its relevance?

Assess on a scale between 0 = *substandard* & 1 = *excellent*

0 I ----- I 1

IIIa: Motor control of voice

Tone (stress level), custom variation and audibility (clarity)?

Assess on a scale between 0 = *substandard* & 1 = *excellent*

0 I ----- I 1

IIIb: General motor performance

How are physical movements performed and the weapon handled, e.g., muzzle point control and use of sights etc.?

Assess with a cross between 0 = *substandard* & 1 = *excellent*

0 I ----- I 1

IIIc: Spatial tactical implementation

Selection of position and movements in the scenario room during the course of the scenario to address the task?

Assess on a scale between 0 = *substandard* & 1 = *excellent*

0 I ----- I 1

V: Temporal tactical implementation

When in time were decisions made and what were their relevance (choice of action)?

Assess on a scale between 0 = *substandard* & 1 = *excellent*

0 I ----- I 1

VI: Handling of the situation overall

Assess on a scale between 0 = *substandard* & 1 = *excellent*

0 I ----- I 1