

The Workload and Performance Relationship in the Real World: A Study of Police Officers in a Field Shooting Exercise

Tal Oron-Gilad

Ben-Gurion University of the Negev, Beer-Sheva, Israel

**James L. Szalma
Shawn C. Stafford
Peter A. Hancock**

University of Central Florida, Orlando, USA

We examined the relationship between perceived workload and performance by evaluating the responses of police officers to 4 different draw-and-shoot tasks in a night field training exercise which was part of their regular training regimen. Sixty-two police officers volunteered to participate. Results demonstrated an associative trend among 3 tasks where shooting performance decreased and workload increased as the tasks became more complex. However, performance on 1 specific shooting task did not correlate with any of the other 3 tasks, and in this 1 exceptional case, insensitivities were observed in which workload increased but performance remained constant.

field study police officers shooting training workload performance

1. INTRODUCTION

Many of the methods and tools of the ergonomist are developed and refined in laboratory-based evaluations. However, the purpose of these tools is to provide insight and information as to performance response in the actual workplace. Due to the persistent and unfortunate divide between the academic and professional worlds, frequently the academic conceives of and creates such measurement instruments but the professional has to adapt and apply them in the much more challenging circumstances of everyday work.

Although this disconnect can cause frustration, it is valuable to understand and explore the utility of scientific assessment methods in circumstances much less pristine than the research laboratory. However, relatively few academic researchers take on this challenge.

One important methodology that impacts much of work efficiency is that of mental or cognitive workload assessment. Traditional methods of measuring work, such as time-and-motion approaches, examine the objective performance response of the individual since output is often the primary, and sometimes the only, concern of

This work was supported by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office under grant DAAD19-01-1-0621, P.A. Hancock, Principal Investigator. The authors wish to thank Dr. Elmar Schmeisser, for providing administrative and technical direction for the grant. The views expressed in this work are those of the authors and do not necessarily reflect official Army policy. The authors would also like to express their gratitude to the 71 police officers who volunteered to participate for the sole purpose of contributing to a research study. A. Greenwood-Ericksen, J.F. Morgan, and J.M. Ross, who assisted in data collection and reduction, are also acknowledged.

Correspondence and requests for offprints should be sent to Tal Oron-Gilad, Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, P.O. Box 653 Beer-Sheva, 84105, Israel. E-mail: <orontal@bgu.ac.il>.

management. However, the associated cognitive costs of maintaining or improving response capacity may not be transparent to well-trusted approaches such as statistical process control. Workers may be able to maintain their output only at the expense of additional cognitive effort which will not be recorded or elucidated by such approaches which evaluate only overt behavior. Thus, the field assessment of cognitive workload is a critical issue and one that we seek to address in the present work.

The particular circumstances investigated here are a police night-time firearms training exercise which was part of the routine training program of police officers in the south-eastern part of the USA. Such training procedures are intended to both maintain and improve officers' shooting skills. The primary goal for this study was to examine the contribution of adding subjective workload measurement to the objective performance measures (e.g., shooting accuracy and time on task) that are already obtained on a routine basis. Thus, our goal was to examine whether the measurement of subjective workload can add information to the training officer that cannot be directly attained from the objective performance measures. Better understanding of the officers' performance and workload relationship can lead to improved training procedures and thereby enhance shooting performance in real-world threatening situations.

Primary task performance is often considered the major reflection of mental workload [1], but here we take it as the criterion measure against which other subjective measures are necessarily compared. If the various measures of subjective workload always followed performance (i.e., a deterministic *association*) [2, 3, 4] such measures would have little or no informational value. Similarly, if there was a complete and persistent *dissociation* between a subjective measure of workload and performance, the former measure would have little practical utility. In some performance tasks, such as vigilance, reports have recorded a direct and consistent association between performance and subjective mental workload (e.g., Warm, Dember, and Hancock [5]; but see Szalma, Warm, Matthews, et al. [6]).

In these situations, increases in task difficulty simultaneously induce both performance decrement and an increase in subjective workload. In contrast, investigators examining multiple-task conditions (e.g., Yeh and Wickens [4]) have reported dissociation between perceived workload and performance.

In an attempt to distinguish the formal link between workload and performance, Yeh and Wickens [4] provided an approach using a framework that is grounded in attentional resource theory [7, 8, 9]. They proposed that dissociation between performance and workload occurs under a number of circumstances that include the following: (a) when greater resources are invested to improve resource-limited tasks (cf., Norman and Bobrow [10]); (b) if demands on working memory are increased by time-sharing; and (c) when performance is sensitive to some subtask element while subjective measures reflect more global demands. Dissociations can also occur when greater resource investment is induced through increasing motivational incentives (e.g., Vidulich and Wickens [11]). In addition to association and dissociation, insensitivities have also been observed [2, 3]. Insensitivities represent the cases in which either the primary task performance or the subjective workload changes but the other does not. The full pattern of possible relationships indicates that dissociations have to be distinguished from insensitivities (Figure 1); thus, a case in which performance is stable among conditions but subjective workload increases indicate a tradeoff in which the individual maintains performance only by exerting more effort. By contrast, when performance changes but workload does not, the individual shows that at that moment they may be less aware of their own task response level.

Perceived workload can also reflect of the level of stress that the individual experiences while performing a task [12]. Hancock and Warm's [13] model of adaptation under stress predicts that increased perceived workload precedes any performance change and therefore a systematic pattern of workload–performance insensitivities and dissociations are expected as long as the individual remains within the psychological

workload	increase	association	performance insensitivity	dissociation
	no change	workload insensitivity	control	workload insensitivity
	decrease	dissociation	performance insensitivity	association
		degradation	no change	improvement
		performance		

Figure 1. Matrix of performance and workload association and dissociations. Notes. Adapted from Parasuraman and Hancock [3].

zone of maximal adaptability. This sequence is shown in Figure 2 and is linked to a more formal description in Table 1. Practical investigations, such as the one presented here, are rarely explored and more data on how the psychological models work in a real-world setting are crucially needed.

TABLE 1. Predictions of Workload and Performance Relationships Based on Hancock and Warm’s [13] Extended-U Model

Region	Workload	Performance	Relationship
a	↑↑	↑↑	dissociation
b	no change	no change	control
c	↑↑	no change	performance insensitivity
d	↑↑	↓↓	association
e	plateau 100%	↓↓	workload insensitivity
f	↓↓	↓↓	dissociation

2. EXPERIMENTAL HYPOTHESES

With regard to the relationship between subjective workload and performance we derived four hypotheses.

- For each task separately, officers who are less skilled in handling firearms will experience increased workload relative to those who are more skilled because they devote more attentional resources and effort to compensate for their lack of skill.

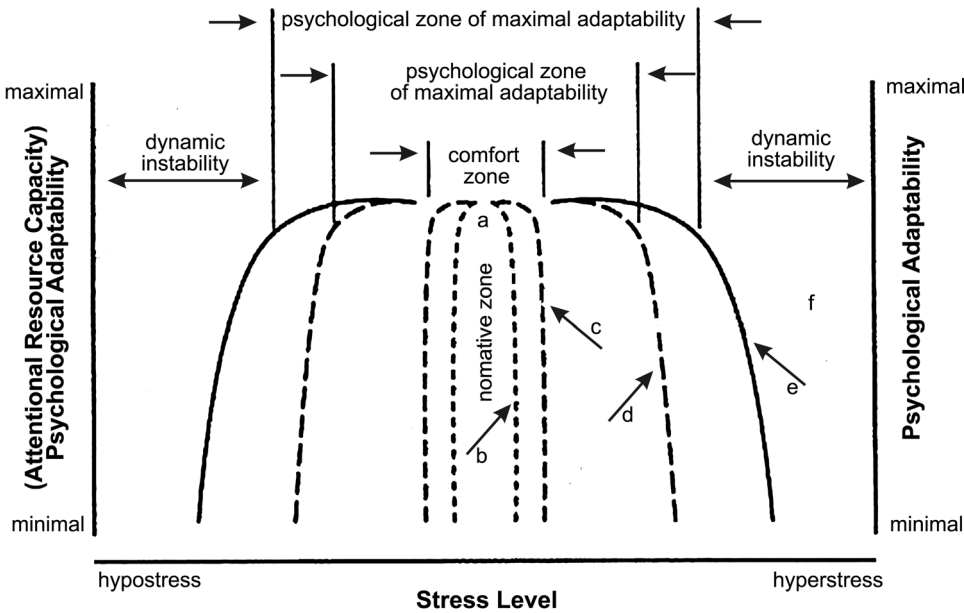


Figure 2. Hancock and Warm’s [13] extended-U model of stress and performance. Notes. See Table 2 for predictions on workload–performance relationships for each region. Task demand and dynamic changes in task demands are the major source of stress in this study; therefore the stress level axis in this model can also be labeled as task demand.

- When comparing across tasks, associations (in which performance decreases and workload increases) or performance insensitivities (in which workload increases but performance remains constant) will occur. Cases in which association occurs indicate that as tasks become more difficult greater effort is exerted to deal with increasing task demand. Furthermore, for skilled officers, devoting additional mental resources to shooting performance might paradoxically have a negative effect on performance, as they “over-think” the task and to some degree “choke” (block normal breathing) [14, 15, 16]. Thus, implicit skills gained through practice can break down when individuals devote additional attentional resources in an attempt to improve performance.
- With regard to the novelty of tasks, automaticity is a fundamental characteristic of high performance skills [17, 18]. Those who are skilled may show performance insensitivities because they can devote their resources to coping with higher demands while protecting their performance. However, those low in shooting skill cannot devote resources to dealing with greater demands without sacrificing performance (association). Alternatively, if they are already exerting maximum effort in lower demanding tasks adding higher demands will impair performance but leave workload unchanged.
- With regard to immediate feedback, providing immediate performance feedback is a traditional method for seeking performance improvement [19]. It may have a significant role in generating dissociations or workload insensitivities by facilitating the development of an internal representation for participants to assess their performance, and by increasing awareness of suboptimal performance [20]. In this study, only one of the novel shooting tasks (the metal targets) provided immediate feedback regarding shooting accuracy.

3. NIGHT SHOOTING EXERCISE

Night training exercises are not obligatory in the police training regimen in the south-eastern USA. The particular police department has voluntarily been conducting night shooting exercises as part of their officers training program. In this case, it is the training officers' responsibility to design the exercise and the shooting task manipulations based on their experience and understanding. Furthermore, as a discretionary exercise, there are no official instructions or qualifying (pass/fail) requirements.

4. EXPERIMENTAL METHOD

4.1 Experimental Participants

All police officers were required to participate in the night shooting exercise. However, participation in our research component was voluntary and each individual officer was free to decide whether or not to participate. Of the 91 officers who completed the exercise, 71 (78%) volunteered to participate in our research. In some cases officers did not complete the questionnaires between shooting tasks as requested. This left a total of 62 participants who completed the entire data series and it was these data which were subjected to analysis. These included 10 women and 52 men, who had a mean age of 37 years (range 22–56), with an average of 11 years of police experience (range 1–32).

4.2. Firearms

Each officer used a SIG (Schweizerische Industrie Gesellschaft™, Switzerland) Sauer p226 9-mm handgun. All officers used the same standard issue duty belt and equipment. All officers also carried a standard department issued flashlight. All officers were required to wear sound attenuating hearing protection (external muffs or internal plugs), body armor, and clear eye protection during each shooting task.

4.3.1. Task design

The night exercise included four different shooting tasks. Those were (a) the warmup task, (b) the flashlight task, (c) the barrel task, and (d) the metal task. The warmup, flashlight, and barrel tasks were held in the same shooting environment. The metal task had different requirements as we detail in section 4.3.5. Warmup and flashlight were group exercises where teams of officers participated at the same time under the overall direction of the training officer. All officers were familiar with the warmup task and with the use of a flashlight while shooting (flashlight task) from their previous training exercises. The barrel and metal tasks were introduced for the first time in this exercise and were performed by each officer separately. The tasks were predesigned by the training officers so there was no opportunity for any direct experimental control. The only research intervention permitted was that the order of the two more complex tasks (barrel and metal tasks) varied among participants; hence, the order of the tasks was warmup, flashlight, barrel, and metal, or warmup, flashlight, metal, and barrel. In

keeping with the noninterventionist nature of this study we had no control as to who was assigned to either of the orders and therefore we cannot claim homogeneity of the number of participants in each group or homogeneity of variance in performance ability.

4.3.2. Warmup task

This was a basic draw-and-shoot task, in which each officer was required to fire a total of 16 rounds at a standard paper target of a blue silhouette of a human figure against a white background. All rounds were fired in very low ambient light conditions in which the blue silhouette of the target was barely visible (Figure 3). Performance on this task was evaluated by the percentage of hits out of the 16 shots fired. A hit was defined as a shot placement consistent with a lethal or incapacitating wound, and a miss was defined as a shot placement outside the boundaries of the target silhouette or was inconsistent with a mortal or incapacitating wound (e.g., a hit in the arm of the target figure).



Figure 3. Police officers shooting at paper silhouettes (warmup task).

4.3.3. Flashlight task

This task consisted of a 24-round shooting exercise with reloads, while the flashlight was held and body positions rotated. Some portions of the task required officers to fire their weapons in very low ambient illumination conditions, whereas at other points officers were required to illuminate targets with their flashlights. Performance was evaluated by the percentage of hits out of the 24 shots fired.

4.3.4. Barrel task

This task was performed individually by each officer. Five paper silhouette targets were positioned along the shooting range target lane and 5 barrels were positioned accordingly across the shooting range fire lane. The barrels were equally distributed over 27 m (~6 m between two adjacent barrels). The officer was required to move sequentially from barrel to barrel and shoot three rounds as quickly as possible at the appropriate target. After engaging a target, the officer ran to the next barrel to engage the following target. For barrels 1, 2, and 5, the officer reached the barrel and fired immediately. For barrels 3 and 4, the officer was instructed to take cover behind the barrel and illuminate the target with a flashlight prior to engaging it. Performance was evaluated by the percentage of hits out of the 15 shots fired and by the time required for each officer to complete the entire sequence.

4.3.5. Metal task

This task was also performed individually by each officer and was conducted in a different area of the shooting range illuminated by the strobe lights of a police cruiser. The officer was required to aim at square metal targets painted in three different colors distributed across two firing lines at two distances. The officer was required to aim at one of three (for the first shooting line) or four (for the second shooting line) possible targets. The task began at the first firing line. As the officer pointed the weapon toward the metal targets, the instructor called out a color. The officer had to hit the metal target of that color

twice consecutively. Thus, the officer fired as many rounds as necessary until two consecutive hits were made. This procedure was repeated twice (a total of 4 hits required). Once completed, the officer proceeded to the second firing line. Then the instructor called out a second color and the officer had again to shoot and hit the target twice consecutively. This procedure was also repeated twice (4 hits altogether). Unlike the paper silhouettes, the metal targets provided auditory feedback after each shot as a result of the sound of the bullet hitting (or missing) the metal target (the hearing protection provided attenuation of the noise but did not completely mask the sound of the bullet hitting the target). Furthermore, the metal targets were smaller than the paper ones and the officers were not familiar with using them in firearms training. Performance on this task was evaluated by the time required for each officer to complete the entire sequence and by the percentage of hits out of the total shots fired. Note that in this task the number of hits was constant ($4 + 4 = 8$) while the number of possible shots fired varied among officers.

4.3.6. Secondary task: time estimation

During the barrel and metal tasks officers were required to estimate task duration immediately following their performance. Prospective duration estimation (in which the person is aware that a duration judgment will be made at the end of a task) can serve as an effective secondary task for workload measurement [21]. Actual (clock) time and estimated time scores were used to compute the duration judgment ratio (DJR) [21], which is the ratio of the estimated time and the actual time in minutes and seconds, expressed as a percentage.

4.3.7. Subjective workload assessment

Mental workload was assessed using the raw NASA task load index (RTLX) scores, an unweighted average of the subscale values. In the original version of the TLX [22], paired comparisons were used to derive weights for the six subscales of the TLX. However, as Byers, Bittner, and Hill [23] showed and as Nygren

[24] also discussed at length, RTLX scores can provide an even better account of the workload experienced by the participant than traditional weighted TLX scores. The RTLX is composed of six sources of workload: mental demand, physical demand, temporal demand, frustration, effort, and self-rated performance. Following each task, officers rated their perceived workload on these individual sources, each on a 0–100 scale. Global RTLX estimates were derived by calculating the average of the subscale values.

4.4. Experimental Procedure

The exercise occurred in January 2004 at an outdoor police shooting range in the south-eastern part of the USA. The average temperature at this time was approximately 10 °C with no rain. The exercise was performed in darkness (after 7 p.m.), and the total session time varied from 60 to 120 min depending on the number of officers in a session. Officers were trained in groups of 6–12 at a time. Prior to the beginning of the exercise the officers were pre-briefed by the training officer on safety precautions and on task composition. The pre-briefing was held in a well lit lecture room, which was part of the shooting range facility. At the end of the pre-briefing the officers were briefed on the research and its purpose and then asked if they would volunteer to participate. The experimenter then presented the RTLX questionnaire and briefly explained its components. Officers were asked to complete the questionnaire at the end of each task. The forms were distributed and completed in a lighted area behind the shooting range where the officers loaded their weapons throughout the exercise. Additionally, the officers were told that they would be asked by the experimenter to provide an estimate of the task length (prospective duration estimation) in minutes and seconds upon completion of the barrel and metal tasks. A paper-based version of the RTLX questionnaire was filled after completion of each task and time estimates were collected upon completion of the barrel and metal tasks.

5. RESULTS

5.1. Overall Performance

Shooting performance was assessed separately for each task by calculating the percentage of hits from the total shots made. This measure is the simplest and most direct reflection of shooting accuracy. For the barrel and metal tasks, which were performed individually by each officer, task duration data was collected and therefore speed–accuracy tradeoff was examined.

5.1.1. Accuracy

The mean percentages of hits were 76 (*SD* 19), 66 (*SD* 22), 58 (*SD* 21), and 62 (*SD* 15) for the warmup, flashlight, barrel, and metal tasks respectively. A repeated measures analysis of variance (ANOVA) on percentage of hits revealed a significant effect for task ($F(3,183) = 13.54$, $p < .0001$). Post hoc (Tukey's HSD) analyses indicated significant ($p < .05$) differences for hits between the warmup task and the other three tasks, and between the flashlight and barrel tasks.

No order effects were found for changing the sequencing of the tasks between the metal and barrel tasks. Years of service and age were not significant moderators of performance accuracy ($p > .05$ in each case).

Significant correlations were observed for hits between the warmup and flashlight tasks ($r = .58$, $p < .01$), between the warmup and barrel tasks ($r = .47$, $p < .01$), and between the flashlight and barrel tasks ($r = .45$, $p < .01$). However, accuracy on the metal task did not correlate with any one of the other three tasks ($p > .09$ in each case).

5.1.2. Speed–accuracy tradeoffs

Mean duration for the barrel and metal tasks was 59 s (*SD* 13) and 57 s (*SD* 19) respectively. For the barrel task there was no significant relationship between the duration of the task and percentage of hits ($b = -.19$, *SD* .19, $\beta = -.12$, $p > .3$). Thus, those who performed the task faster were not more accurate in shooting or vice versa. This is not surprising because the duration of the barrel task was mostly influenced by the speed of

movement from one barrel to the next and not by shooting accuracy.

For the metal task there was a significant negative relationship between task duration and percentage of hits ($b = -.28$, $SD .92$, $\beta = .346$, $p < .003$), indicating that the better performers also tended to perform the task faster. This finding is also not surprising because the duration of this task was strongly linked with shooting performance. The more accurate shooters fired fewer rounds at each target and therefore required less time to complete the task.

5.2. Workload Estimates

Perceived global workload scores were derived from the (unweighted) average of the six subscale ratings. The mean global workload scores were 43 ($SD 15$), 49 ($SD 16$), 61 ($SD 16$), and 57 ($SD 17$) for the warmup, flashlight, barrel, and metal tasks respectively. A repeated measures ANOVA on these data revealed a significant task effect, $F(3, 183) = 56.27$, $p < .0001$. Post hoc (Tukey's HSD) analyses indicated significant differences in global workload between the warmup task and the other three tasks, between the flashlight and barrel tasks, and between the flashlight and metal tasks ($p > .05$ in each case). No significant order

effects were observed for the metal and barrel tasks. As in the case of performance, years of service were not a significant moderator of global workload or subscale ratings ($p > .05$ in each case).

Table 2 provides the descriptive statistics of the subscales and the F tests for the main effects for the NASA TLX [22] within each task. The physical demand subscale is the only scale where significant differences occurred among all four tasks. There were no other significant differences in subscale ratings between the barrel and metal tasks.

5.3. Workload–Performance Relationship Within Each Task

We hypothesized that officers who were less skilled in shooting would experience increased workload relative to those who were more skilled. Figure 4 provides the scatter plots of workload as a function of performance for each of the four tasks. Although shallow trend-lines ($>-.17$) were observed, simple regression analyses indicated no significant relationship between perceived workload and performance ($p > .1$ in each case).

TABLE 2. NASA Task Load Index (TLX) [22] Subscale Descriptives and Main Effects for the 4 Tasks

TLX Element	Task ($M \pm SD$)				Main Effect and Post Hoc (Tukey's HSD)*
	Warmup	Flashlight	Barrel	Metal	
Mental	52 \pm 26	59 \pm 23	68 \pm 21	65 \pm 24	$F(3, 183) = 14.41$, $p < .00001$ W<>B, M; F<>B
Physical	28 \pm 23	40 \pm 26	62 \pm 24	53 \pm 25	$F(3, 183) = 60.07$, $p < 0.0001$ W<>F, B, M; F<>B, M; B<>M
Temporal	38 \pm 26	45 \pm 26	67 \pm 22	62 \pm 25	$F(3, 183) = 50.75$, $p < 0.0001$ W<>F, B, M; F<>B, M
Own performance	63 \pm 24	61 \pm 23	55 \pm 23	52 \pm 22	$F(3, 183) = 4.31$, $p < .01000$ W<>M
Effort	48 \pm 23	53 \pm 24	65 \pm 20	62 \pm 21	$F(3, 183) = 17.70$, $p < .00001$ W<>B, M; F<>B, M
Frustration	29 \pm 23	38 \pm 23	47 \pm 27	49 \pm 25	$F(3, 183) = 16.95$, $p < .00001$ W<>F, B, M; F<>B, M
Global RTLX	43 \pm 15	48 \pm 17	57 \pm 16	57 \pm 18	$F(3, 183) = 56.27$, $p < .00001$ W<>F, B, M; F<>B, M

Notes. * $p < .05$, $N = 62$. When calculating the global raw NASA TLX (RTLX) the performance subscale is reversed. W—warmup, F—flashlight, B—barrel, M—metal.

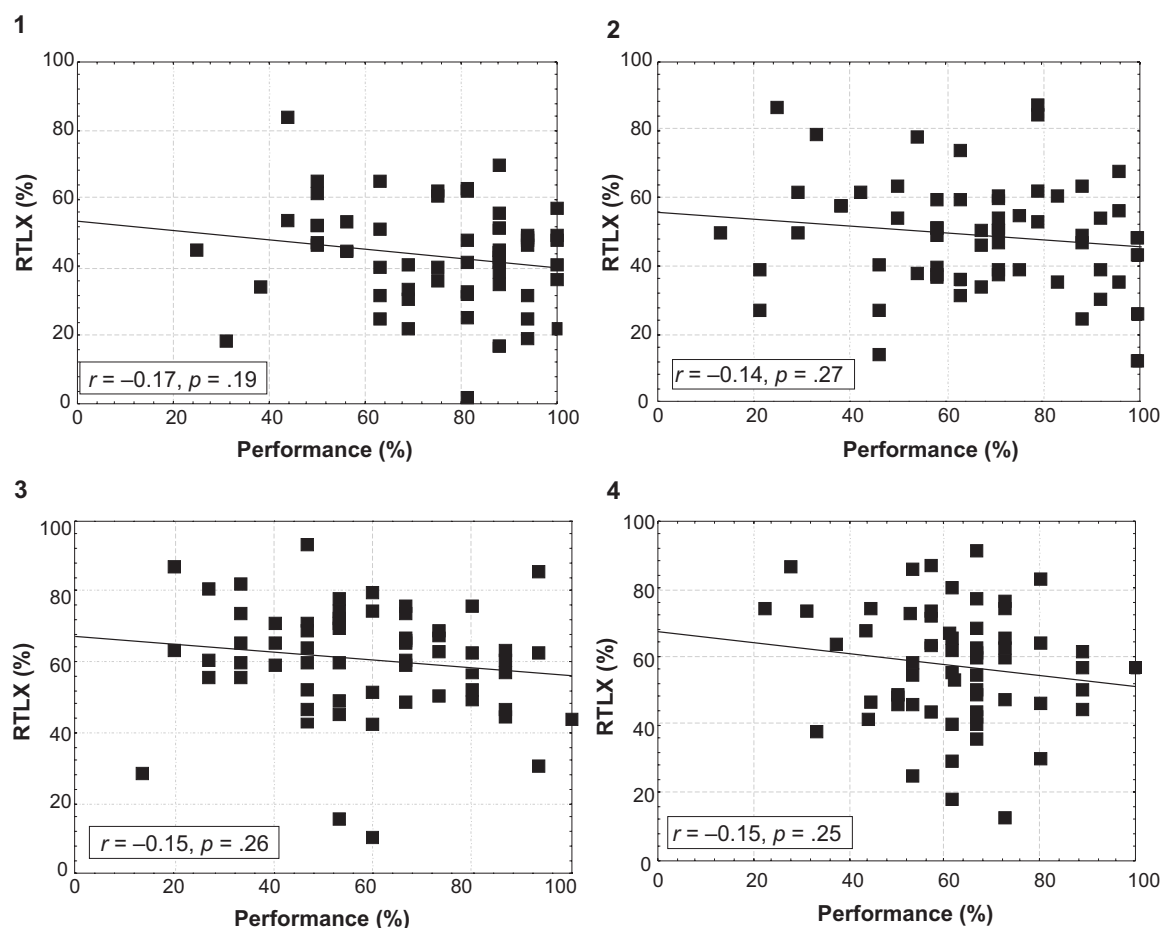


Figure 4. Scatter plots of workload (RTLX) scores and performance scores for each task. *Notes.* RTLX—raw NASA task load index.

5.4. Workload–Performance Relationship

The pattern of workload–performance relationship that has been identified from the performance and global workload estimates is one of both associations and insensitivities. Overall, across the tasks, there was an associative trend (moving from higher to lower performance was accompanied by a corresponding increase in workload), as shown in Figure 5. A repeated measures ANOVA on Task (4) \times rating (performance, workload) was conducted. There was a significant main effect for rating, $F(1, 61) = 20.271, p < 0.0001$, and for the task \times rating interaction, $F(3, 183) = 48.181, p < 0.000001$. Post hoc (Tukey's HSD) tests revealed that there was a significant difference between performance and workload for the warmup and the flashlight tasks but not for the barrel and metal tasks.

Overall the barrel and metal tasks appeared to be equivalent in performance and workload. However, they differed in their relation to the warmup and flashlight tasks. The barrel task had a significant correlation in performance with the warmup and the flashlight tasks. For those three tasks a significant increase in workload and a comparable significant decrease in performance were observed. The associative trend among these three tasks is therefore evident. As the tasks became more complex, officers exerted more effort but performance decreased up to the point at which workload and performance were equivalent (i.e., not significantly different) in the barrel task. For the metal task a different trend was observed, such that workload was significantly higher for the metal task relative to the flashlight task but performance remained

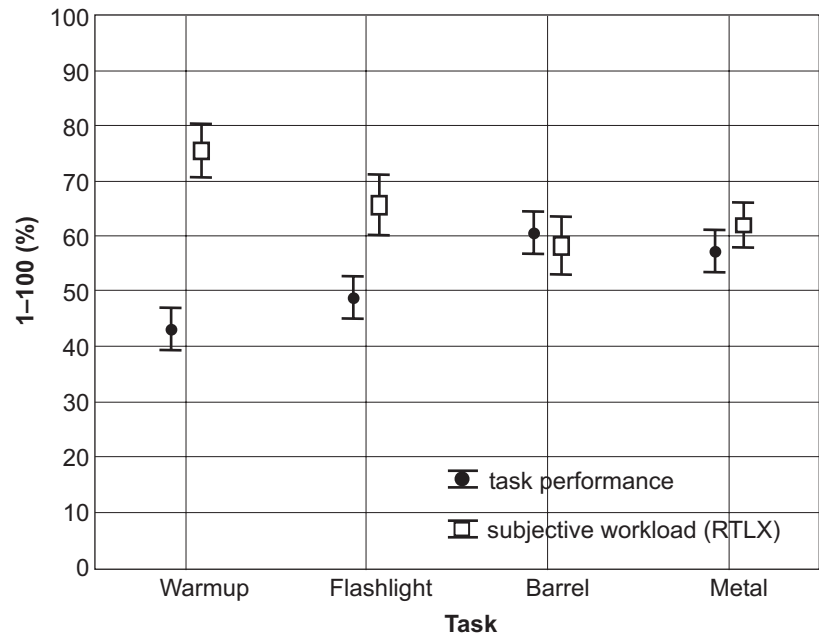


Figure 5. Workload and performance relationship among the 4 tasks. Notes. RTLX—raw NASA task load index.

equivalent for the two conditions (i.e., not significantly different). Hence, more effort was exerted but no significant change in performance occurred, a pattern of performance insensitivity.

To further investigate the source of difference between the two more complex tasks, a hierarchical regression model of workload and performance was computed. This model was not intended to predict performance but rather to explain which workload dimensions affected (or were affected by) performance and the strength of these effects (see Cohen, Cohen, West, et al. [25]). For the barrel task, a hierarchical, linear regression was performed on shooting performance with shooting skill (defined here as shooting performance in the warmup task) entered first, followed by each TLX subscale. This was done so that regressions could be compared to determine whether shooting performance in the barrel task was mediated by the initial shooting skills while examining the relationship between the six individual TLX subscales and shooting performance. The only model where a subscale added significantly to the prediction of the shooting performance was the regression model with the mental demand subscale: $b_{\text{shooting skill}} = .48$ ($SE_b = .11$), $\beta = .44$, $p < .0001$,

$R^2 = .22$ and $b_{\text{mental subscale}} = -.25$ ($SE_b = .10$), $\beta = -.26$, $p < .05$, $\Delta R^2 = .07$ for step 2 ($ps < .05$).

A similar analysis was conducted for the metal task. In this case overall performance did not correlate with the warmup task, indicating that shooting skill did not influence performance in the metal task. To examine the relationship between the six individual TLX subscales and the metal task shooting performance, a series of hierarchical linear regressions were performed on shooting performance with the individual TLX subscales. Only the models in which frustration and own performance were entered resulted in a significant regression. The model that accounted for the most variability in shooting accuracy was one that combined both subscales: $b_{\text{frustration subscale}} = .18$ ($SE_b = .06$), $\beta = -.32$, $p < .005$, $R^2 = .14$ and $b_{\text{performance subscale}} = -.23$ ($SE_b = .07$), $\beta = -.34$, $p < .005$, $\Delta R^2 = .12$ for step 2 ($ps < .005$). Note that there was no statistically significant correlation between these two subscales, $r = .05$.

5.5. Time Estimation

Prospective time estimations were made for the barrel and metal tasks. The mean estimated time

was 64 s (*SD* 44) and 66 s (*SD* 51) for the barrel and metal tasks respectively. DJR for the barrel and metal tasks was 109 (*SD* 68) and 116 (*SD* 70) respectively. The correlation between DJR for the two tasks was significant ($r = .456, p < .001$). Correlations between time perception, and performance and workload were not statistically significant ($p > .05$ in each case).

6. DISCUSSION

This study provided initial evidence that a significant link between performance and workload does not necessarily occur. Poor performance was not always associated with higher workload, or vice versa. Hence, as suggested by Vidulich and Wickens [11], resource investment may result from increasing motivational incentives or individual differences among officers, so that workload is not deterministically tied to the quality of performance.

In this study a pattern of association and performance insensitivities was also confirmed, with an associative trend among the warmup, flashlight, and barrel tasks. These results indicated that as the task became more complex performance decreased significantly, even as officers exerted significantly more effort to perform well. Performance insensitivities were observed for the metal task; performance did not vary significantly between the flashlight and metal tasks but perceived workload did. In terms of Hancock and Warm's [13] model, the officers were operating within the threshold of the comfort zone (moderate level of stress, see zone b in Figure 2) where performance insensitivities were predicted to occur. Furthermore, only under extreme conditions would one expect time estimates to be predictive of the stress state of the individual [26]. Such findings attest to the adaptability of the human operator; even though the performed tasks were complex and required co-ordinated mental and physical work under time pressure in a suboptimal environmental condition. A likely reason the officers remained within the thresholds of Hancock and Warm's comfort zone is that training exercises like

the ones presented here were not intended to physically challenge the participants.

The insensitivities observed in this study can be examined from the perspective of the multidimensionality of workload as reflected in the NASA TLX [22] subscales. For this purpose we investigated in more detail the performance insensitivities we found for the two more complex tasks (barrel and metal). This analysis showed that some dimensions of the global workload contributed more than others to the workload assessments of shooting tasks. For the metal task, the self-rating subscales were the better markers for accuracy. Specifically, the frustration and own-performance subscales significantly correlated with shooting accuracy. By contrast, the mental demand subscale combined with basic shooting skill were markers for performance accuracy in the barrel task. Given the physical nature of the task, which involved running and taking cover while engaging targets, one might have expected that the effort and physical demand subscales would have exerted substantial influences. Furthermore, researchers often tend to underestimate the ability of individuals to estimate their own performance (see Johnson [20]). However, in the metal task such underestimation would be unjustified because the own-performance estimations accounted for a significant portion of variability in performance accuracy. Possibly, the immediate feedback provided in the metal task made it possible for officers to more accurately estimate their performance and to attempt to improve it throughout the task (see also Newell [19]). On the other hand, this immediate feedback may have also generated more frustration directly related to task performance. For the barrel task, the finding that mental demand accounted for a significant proportion of variance in shooting performance indicated that officers who expended the most mental effort were also the poorest performers. This represents a workload–performance association.

Although it is impossible to infer from the current dataset the cause of the differences between the metal task and the three other tasks, it is clear that providing officers with tasks other

than traditional stand-and-fire shooting range tasks require additional skills and thereby have the potential to enhance training outcomes. Furthermore, these results underscore the importance of examining the subscales of the TLX and confirm that the TLX subscales have diagnostic value, an important criterion for any workload measure [27].

Finally, for the metal and barrel tasks the current study has demonstrated that dissociations and insensitivities are more likely to occur when novel tasks are introduced into a training regimen because individuals may not possess sufficient skills to perform even if they exert the additional effort. In sum, we have shown that investigation of the workload–performance relationship in a real-world setting can be informative and diagnostic, and we have demonstrated how such analyses may be done to better understand the relationships between these variables.

7. SUMMARY AND CONCLUSIONS

The shooting exercise consisted of four tasks. Three of them required aiming at paper targets with human silhouettes. Each of the three tasks was progressively more difficult both in physical demand and co-ordination. The increase in difficulty was manifested in substantial changes in shooting performance and in the perceived workload ratings. Hence, shooting performance decreased and workload increased as the tasks became more complex. Yet, there was a significant correlation between the tasks indicating that the officers' performance on the basic draw-and-shoot task (warmup task) was predictive of their performance in the other two more demanding tasks, implying perhaps that all three tasks were tapping the same shooting skills or that shooting skill serves to protect performance when greater complexity is introduced. Performance on the metal task did not correlate with either of the less difficult paper-target tasks (warmup or flashlight) but yielded similar workload to the most difficult paper-target task (barrel). This task consisted of a different shape of metal targets which provided an immediate feedback to the officer as to whether

or not the target had been hit. Finally, this task also required rapid decision making. At each stage of the task the officer was required to shoot either one out of three or one out of four potential targets depending on the color of the target as instructed by the training officer. The source of insensitivity can be attributed to the novelty of the task, to the immediate feedback, and perhaps to the higher cognitive demand required in the metal task (derived from listening to instructions while aiming at targets).

This study is an initial exploration of this issue; its results indicate that measurement of performance and workload showed a pattern of association and insensitivity that is diagnostic. This indicates the need to expand police training with firearms beyond the traditional stand-and-shoot procedures. Designing tasks that impose additional real-world demand (e.g., holding a flashlight, running, and aiming at more than one possible target at a time) may have beneficial effects in improving performance by forcing officers to practice shooting skills in context. Further research is needed to identify the task demands to be added to optimize training outcomes.

REFERENCES

1. Hancock PA, Meshkati N, editors. Human mental workload. Amsterdam, The Netherlands: North-Holland; 1988.
2. Hancock PA. Effects of control order, augmented feedback, input device and practice on tracking performance and perceived workload. *Ergonomics*. 1996;39:1146–62.
3. Parasuraman R, Hancock PA. Adaptive control of central workload. In: Hancock PA, Desmond PA, editors. Stress, workload, and fatigue. 1st ed. Mahwah, NJ, USA: Erlbaum; 2001. p. 305–20.
4. Yeh YY, Wickens CD. Dissociations of performance and subjective measures of workload. *Hum Factors*. 1988;30:111–20.
5. Warm JS, Dember WN, Hancock PA. Vigilance and workload in automated systems. In: Parasuraman R, Mouloua M, editors. Automation and human performance. Mahwah, NJ, USA: Erlbaum; 1996. p. 183–200.

6. Szalma JL, Warm JS, Matthews G, Dember WN, Weiler EM, Meier A, et al. Effects of sensory modality and task duration on performance, workload, and stress in sustained attention. *Hum Factors*. 2004;46:219–33.
7. Kahneman D. *Attention and effort*. Englewood Cliffs, NJ, USA: Prentice-Hall; 1973.
8. Gopher D, Donchin E. Workload: an examination of the concept. In: Boff KR, Kaufman L, Thomas JP, editors. *Handbook of human performance*. Vol. 2. Cognitive processes and performance. New York, NY, USA: Wiley; 1986. p. 41.1–41.49.
9. Wickens CD. The structure of attentional resources. In: Nickerson R, editor. *Attention and performance VIII*. Hillsdale, NJ: Erlbaum; 1980. p. 239–57.
10. Norman DA, Bobrow DG. On data-limited and resource limited processes. *Cognit Psychol*. 1975;7:44–64.
11. Vidulich MA, Wickens CD. Causes of dissociation between subjective workload measures and performance. Caveats for the use of subjective assessments. *App Ergon*. 1986;17:291–6.
12. Hancock PA, Desmond PA, editors. *Stress, workload and fatigue*. 1st ed. Mahwah, NJ, USA: Erlbaum; 2001.
13. Hancock PA, Warm JS. A dynamic model of stress and sustained attention. *Hum Factors*. 1989;31:519–37.
14. Beilock SL, Carr TH. On the fragility of skilled performance: What governs choking under pressure? *J Exp Psychol Gen*. 2001; 130:701–25.
15. Beilock SL, Carr TH, MacMahon C, Starkes JL. When paying attention becomes counterproductive: Impact of divided versus skill-focused attention on novice and experienced performance of sensorimotor skills. *J Exp Psychol Appl*. 2002;8:6–16.
16. Masters RSW. Knowledge, knerves, and know-how: the role of explicit versus implicit knowledge in the breakdown on a complex motor skill under pressure. *British Journal of Psychology*. 1992;83:343–58.
17. Schneider W, Fisk AD. Attentional theory and mechanisms for skilled performance. In: Magill RA, editor. *Memory and control of action*. Amsterdam, The Netherlands: North-Holland; 1983. p. 119–43.
18. Schneider W, Shiffrin, RM. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychol Rev*. 1977;84(1):1–66.
19. Newell A. The knowledge level [Presidential address]. *AI Magazine*. 1980;2(2):1–20.
20. Johnson A. We learn from our mistakes—don't we? *Ergon Des*. 2004;12:24–7.
21. Block RA, Zakay D, Hancock PA. Developmental changes in human duration judgments: a meta-analytic review. *Dev Rev*. 1999;19:183–211.
22. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: P.A. Hancock and N. Meshkati (Eds.), *Human mental workload*. Amsterdam, The Netherlands: North-Holland; 1988. p. 139–83.
23. Byers JC, Bittner AC, Hill, SG. Traditional and raw task load index (TLX) correlations: are paired comparisons necessary? In: Mital A, editor. *Advances in industrial ergonomics and safety, I*. London, UK: Taylor & Francis; 1989. p. 481–5.
24. Nygren TE. Psychometric properties of subjective workload measurement techniques: implications for their use in the assessment of perceived mental workload. *Hum Factors*. 1991;33:17–33.
25. Cohen J, Cohen P, West SG, Aiken LS. *Multiple regression/correlation analysis for the behavioral sciences*. 3rd ed. Mahwah, NJ, USA: Erlbaum; 2004.
26. Hancock PA, Weaver JL. Temporal distortions under extreme stress. *TIES*. 2005;6(2): 193–211.
27. O'Donnell RD, Eggemeier FT. Workload assessment methodology. In: Boff KR, Kaufman L, Thomas JP, editors. Vol. 2. *Cognitive processes and performance*. New York, NY, USA: Wiley; 1986. p. 42.1–42.49.