Effect of Workload History on Task Performance

Luz-Eugenia Cox-Fuenzalida, University of Oklahoma, Norman, Oklahoma

Objective: This study investigated the effects of workload history (specifically, sudden shifts in workload) on performance. **Background:** In 1993 the National Research Council identified workload transition as an important concern for human factors researchers. The study of workload history suggests that what an individual has been doing prior to a point in time has an effect on subsequent performance. One trend emerging from workload history studies is that a general decrement in performance is most likely to occur following a decrease in task demand. **Method:** The 198 participants were randomly assigned to a high-to-low or low-to-high condition. Participants performed a version of the Bakan Vigilance Task while correct responses, response times, and total errors were recorded. **Results:** Results supported previous research suggesting a workload decrease results in a performance decrement. More importantly, this study reports that either a sudden increase or decrease could lead to a loss in accuracy and a slowing of response time in a longer time course. **Conclusion:** An explanation of the decrement is offered in terms of adaptation models. In addition, a follow-up study suggested that the decrement is a result of something inherent in the workload shift rather than an effect of fatigue. **Application:** Workload history (more specifically, a workload shift) has significant implications for many work environments. These implications are particularly salient in occupations where individuals are confronted with varying levels of workload demand, especially safety-sensitive occupations.

INTRODUCTION

Most human task performance research has focused on determining task performance capability during a specific time period or exploring the influence of stress, sleep loss, drugs, individual differences, and other variables. Previous research has suggested that the type and duration of work performed prior to a given point in time (i.e., workload history) may have a strong influence on work performance following that point (Cumming & Croft, 1973; Goldberg & Stewart, 1980; M. L. Matthews, 1986). Studies on the effects of workload history on performance have generally been limited to the effects of workload transitions. This important but limited focus has circuitously limited the conceptualization of "workload history," failing to capture the essential, more general, and overarching potential effects of a number of workload history variations on performance.

However, experimental studies focusing on cognitive performance have typically kept workload constant, at different levels. This has often been done in an effort to control for possible confounding variables or to explore effects at a stable workload level (M. L. Matthews, 1986). Although this method has enabled the study of individual responses during fixed workload levels, it has not contributed to the understanding of how individuals respond to dynamic workload situations, which tend to be more representative of many real-world jobs. In fact, it is not inconceivable that this attention to the experimental control of workload (which has led to many advances in understanding the effects of fixed workload) has inadvertently suppressed the study of broader workload dynamics, such as workload history. As the value of predictive efficiency in human performance increases, so does the need for a definition of workload history capable of capturing not only workload variations but also

Address correspondence to Eugenia Cox-Fuenzalida, Department of Psychology, 455 W. Lindsey St., DHT–705, University of Oklahoma, Norman, OK; cox-fuenzalida@ou.edu. *HUMAN FACTORS,* Vol. 49, No. 2, April 2007, pp. 277–291. Copyright © 2007, Human Factors and Ergonomics Society. All rights reserved.

Downloaded from hfs.sagepub.com at UNIV OF OKLAHOMA on January 20, 2016

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/25534

the simplicity inherent in many systems operations and/or sustained workload situations characteristic of other work environments. It is important to broaden the understanding of "workload history" and therefore redefine the conceptualization and increase understanding of its effects on performance.

The following definition of workload history might best account for a wide range of workable theoretical designs in which the effects of workload history are considered. Workload history may be best defined as prior work activity that has an effect on subsequent work activity. In other words, what an individual was doing (precritical work period) prior to a point in time (critical point) has an effect on subsequent performance (postcritical work period). According to this definition, workload history involves a dynamic process potentially reflecting any number of design possibilities.

Although little is known today about the characteristics and dynamics of workload history that are relevant to subsequent work performance, in 1993 the National Research Council identified workload transition as an important concern for human factors researchers and, in so doing, underscored the importance of examining the effects of workload variation on performance. Indeed, workload history (and more specifically, a workload shift) has significant implications for many work environments, particularly those where individuals are confronted with varying levels of workload demand (Huey & Wickens, 1993). For example, consider the bus driver who suddenly enters heavy traffic or the air traffic controller who suddenly has significantly fewer planes to manage within an airspace. Both of these examples demonstrate situations in which an individual's workload history may influence performance during a subsequent critical period. It would therefore seem prudent to have a fuller understanding of workload history and its possible influence on later performance.

The purpose of this project was threefold. Addressing methodological limitations of previous studies, the present study was structured to first verify workload history effects and, second, to test theories of why differential decrements in performance might occur in response to this type of change in workload. Lastly, a follow-up study was designed to determine whether the decrement in performance following a workload transition (found in the original study) was the result of time on task or fatigue. The follow-up study utilized a workload history design requiring participants to work at a fixed workload level following a training-baseline protocol identical to that of the original study.

Decrements in Performance and Workload History

In the area of workload history a small number of studies have examined the consequences of workload variation on performance. One trend emerging from these studies has been the finding that a general decrement in performance is most likely to occur in situations where there is a decrease in task demand. Each of these studies has attempted to explain the nature of the decrement in a different way.

Cumming and Croft (1973) reported the effects of workload history on an auditory performance task. Participants performed the task as the difficulty level of the task was systematically raised and lowered in an alternating fashion. Results suggested a significantly greater performance decrement during the decreasing workload phase of the cycle as compared with the increasing workload phase. Cumming and Croft (1973) operationally defined performance in terms of information transmitted per second, and it appears that they made direct comparisons between performance during increasing and decreasing workload phases of the task.

Cumming and Croft (1973) proposed a complex explanation for this finding. Their explanation began by noting a previously reported relationship between performance effectiveness and stimulus frequency – namely, that people tend to respond faster or more accurately to stimuli presented more frequently than to stimuli presented less frequently. Cumming and Croft (1973) also believed that when workload increased, their participants realized that they could not maintain uniformly high response quality, so they chose to accept higher levels of error as the task difficulty increased. Because it was difficult for their participants to know exactly when the cycles changed and the task began to get easier, the participants continued to accept more errors as they moved into the descending phase of the workload cycle (i.e., a period when the task demand was lower). This acceptance of higher error levels (which was adaptive during the previous phase of task) simply

Downloaded from hfs.sagepub.com at UNIV OF OKLAHOMA on January 20, 2016

compounded the previously known phenomenon in which people respond more poorly when signal frequency decreases unexpectedly. As a result, Cumming and Croft (1973) suggested post hoc that this inability to match expectancies regarding changes in task difficulty explained why their participants showed a much larger performance decrement when they moved to a low workload level after performing at a high workload level, compared with when they experienced an increase in task difficulty following a low workload level.

Goldberg and Stewart (1980) reported a study designed to test whether expectancies were responsible for the decline in task performance when workload moved from high to lower levels. They employed a visual performance task. One condition included a display characteristic that served as a visual cue signaling changes in task demand (i.e., characters appeared on the left of the screen with subsequent characters appearing progressively farther to the right and then back to the left as demand increased and decreased, respectively). The purpose of the cue was to inform participants that either an increase or decrease in task demand was taking place. Even when participants were given this cue, performance was still worse when moving from high to lower workload as compared with moving from low to higher workload, suggesting that it is not a person's inaccurate expectations that are responsible for the effect. Rather, Goldberg and Stewart (1980) suggested post hoc that the decrements following a decrease in work demand resulted from a temporary overload of short-term memory (STM). In other words, they proposed that the rapid presentation of information (at the higher workload level) overloaded STM, resulting in a performance decline. They further hypothesized that the overload persisted (after workload level decreased or presentation rate declined) until the rate had been sufficiently reduced and STM was no longer overloaded.

M. L. Matthews (1986) conducted a set of studies to further explore the effects of workload history on visual task performance. His results replicated the findings of Cumming and Croft (1973) and Goldberg and Stewart (1980): Higher decrements in performance result when workload decreases as compared with when workload increases. In addition, M. L. Matthews (1986) tested the theory that the performance decrement was attributable to a failure of STM. He used a visual performance task that did not require the preservation of serial order information (i.e., STM). Participants were required to examine matrix displays containing 3, 6, 9, or 12 target/noise strings for the occurrence of target signals. Target signals consisted of arithmetic expressions (e.g., 22+12<18) that had to be discriminated from nonarithmetic expressions (e.g., 27 – #6 > Y$). Participants had to first locate and indicate the position of a target signal (if present) by pressing an appropriate key and then evaluate and indicate whether the expression was true or false by pressing another key. A blank screen was presented immediately after a participant's response, producing a 3.5 to 6.0 s delay between stimuli. This made carryover effects unlikely from one stimulus to the next, given the known 333-ms limit of visual iconic memory (Hunt & Ellis, 2004). This methodology provided conditions in which memory capacity and demands were not considered significant contributors.

M. L. Matthews (1986) included one condition during which a high workload level suddenly decreased to a low level. Another condition included a low workload level that suddenly shifted to a high level. A significant decrement in performance was found up to 1 min following a sudden workload reduction, but not following a sudden workload increase. Thus, the data failed to support an explanation of the performance decrement based on the STM overload hypothesis provided by Goldberg and Stewart (1980). Instead, M. L. Matthews (1986) proposed that strategic persistence might account for the performance decrement. He proposed that participants were able to mobilize effective strategies when workload increased but then retained these strategies long after the workload level had fallen. For example, when suddenly reduced to a low workload level, participants might have continued the level of effort they applied during the previous period at high workload, thereby overworking or overdriving the task.

This theoretical model might be confused with the explanation of Cumming and Croft (1973). It is important to note that although both discussed strategies, Cumming and Croft (1973) offered an explanation based on an inability to match expectancies. They believed their participants were unable to tell when a decrease in workload had occurred and thus incorrectly "allowed" themselves errors

Downloaded from hfs.sagepub.com at UNIV OF OKLAHOMA on January 20, 2016

at the lower workload level. In contrast, M. L. Matthews (1986) attributed the decrement to overworking or overdriving the task when it was reduced to a lower workload level.

In summary, despite the use of different methodologies, all three studies clearly demonstrated the influence of workload history on subsequent work performance. Specifically, they identified a general decrement in performance (relative to a perfect performance standard) in a situation where an individual moves to a low workload level after working for some period at a higher level. However, upon careful examination of these studies, it appears that design and/or methodological limitations might have compromised data integrity and consequently cast doubt on the generalizability of their findings.

Limitations of Previous Studies

One concern arises from an insufficient number and/or an absence of practice or training trials. The studies of Cumming and Croft (1973) and Goldberg and Stewart (1980) neglected to provide participants with training sessions. This is of special concern, given that a minimum level of training is essential to bring participants to a reasonably stable performance level and consequently minimize contamination of test data from learning effects (Schlegel & Gilliland, 1990). Although M. L. Matthews (1986) provided a training session for all participants, the training session was 40 min long. Furthermore, the characteristics and parameters of the performance task (including workload level) were identical to the testing session. It is problematic that participants were exposed to 40 min of the "treatment effect" during the training session, because the nature and sequencing of the training and testing sessions might have left participants bored or fatigued.

Second, all three studies either failed to collect baseline data or collected what might have been inadequate baseline data for comparison with test data. The studies of Cumming and Croft (1973) and Goldberg and Stewart (1980) collected no baselines. M. L. Matthews (1986) employed a between-subjects design to collect baselines; however, comparisons between baseline and test data were not altogether comparable. Performance from participants engaging in 40 min of training trials involving high-to-low workload transitions followed by 40 min of test trials involving high-to-low workload transitions were compared with baselines collected from other participants engaged in 80 min at a fixed low workload level. Consequently, given the noncomparability of the baseline samples, it was not surprising to find response times during the shift from high to low workload levels that were slower than those for a fixed low workload level.

Furthermore, it should be noted that some work in vigilance research has examined "workload transition" phenomena and may help to inform understanding of workload history effects. For example, Krulewitz, Warm, and Wohl (1975) shifted participants from high-to-low and low-to-high event rates (i.e., they examined changes in performance as a function of the rate at which repetitive and neutral stimuli were presented) and found an improvement in high-to-low postshift performance and a decrement in low-to-high postshift performance. Later, Moroney, Warm, and Dember (1995) reported that "the post-transition performance of observers shifted from a fast-to-slow event rate (high-to-low task demand) remained below that of their continuous slow event rate controls, and was thus unaffected by the shift. In contrast, the post-transition performance of monitors shifted in the opposite direction, slow-to-fast event rate, was affected by the shift" (p. 1375). Thus, although there are a few inconsistencies between vigilance studies examining workload transition and workload history research, findings in the latter generally suggest that moving from one workload to another tends to have detrimental effects on performance. More importantly, both research domains demonstrate that prior work activity has an effect on subsequent work activity.

The primary purpose of the present research project was to verify the effects of workload history on performance. Specifically, after the methodological limitations of previous studies are corrected for, does the decrement in performance following a transition from high to low workload persist? Indeed, until the issues raised in the present study are addressed, the certainty of the negative effect on performance remains questionable. The primary dependent measures examined were indices of performance on the Bakan (1959) Vigilance Task (i.e., correct responses, reaction time, and total errors). These human performance indices have been commonly employed to measure performance decrements in previous

Downloaded from hfs.sagepub.com at UNIV OF OKLAHOMA on January 20, 2016

workload history research (Cumming & Croft, 1973; Goldberg & Stewart, 1980; M. L. Matthews, 1986) and are therefore essential for comparisons with previous research. However, this study also extends its comparison with other areas by examining performance decrements (i.e., breaking down total errors into omission and commission mistakes). These additional measures enable the discernment of information that may be more pertinent to some applied settings than to others. In other words, in certain applied situations the nature of the error is less important than the error itself. However, in other settings, the nature of the mistake may be significantly more or less critical. For example, consider a pharmaceutical prescription-checking task in which a pharmacist routinely checks for labeling errors. In this scenario, the impact of an error of commission (i.e., thinking there is an error in the absence of one) may cost time, whereas an error of omission (i.e., missing an error) may have lethal consequences.

This study had a second purpose as well. Previous studies attempted to explain the nature of the general decrement in performance following a decrease in workload. If the decrement remains, this study was designed to explore the nature of the decrement. The most recent interpretation of the performance decrements was based on strategic persistence (M. L. Matthews, 1986). However, another explanation might be plausible as well. Comprehensive models of adaptation such as the general adaptation syndrome (Selye, 1956, 1978), the state control theory of adaptation (Hockey, 1986, 1997), and the dynamic model of stress adaptation (Hancock & Warm, 1989) would suggest that a decrease in initial performance following high workload levels might be the result of recuperative efforts interfering with task performance (despite the low workload level), as opposed to strategic persistence. Thus, something more akin to an adaptation model might better explain the performance phenomenon following decreases in workload.

Given that the present study was also structured to compare the strategic persistence and adaptation models, a task was selected that allowed a test between these two approaches, based on error rates (i.e., commissions and omissions). If M. L. Matthews (1986) was correct and the participant was maintaining unnecessarily high levels of effort for a low-demand task, then there should be evidence of significant commission errors following the transition from a high to a low level of workload. That is, if the person was overdriving or overworking the task at the low workload level, this may be manifested by an increase in errors of commission. It is indeed possible that the increase in rate of responding at the high workload level may result in a lower decision threshold, and following a sudden decrease in workload it is possible that the lower decision threshold would adversely affect performance. However, if the decrement in performance was attributable to recuperative effort associated with adaptation (because of the cognitive load), then there ought to be evidence of omission errors. In other words, if the person was trying to recover from the high workload level, an increase in omission (or missed signals) should result.

STUDY 1

Method

Participants. Initially, 209 participants were randomly selected and recruited from among approximately 450 lower division psychology students at the University of Oklahoma. Following random selection from overall participant lists, participants for this study were contacted by telephone and invited to participate. Of the individuals contacted, only 5 declined participation. Six participants did not complete the experiment (1 was dropped after reporting hearing subliminal messages, 2 were dropped because of equipment failure, and 3 were dropped for failing to understand the task), leaving a total of 198 participants (106 women and 92 men). Participants received extra credit as one option for fulfilling the requirements of undergraduate psychology courses.

Materials. A computer-based version of the Bakan (1959) Vigilance Task was employed. This is an auditory vigilance task consisting of a series of digits presented to the participants via earphones. Each training and test trial was 3 min in duration. During each trial, in this version of the task participants were instructed to detect odd-even-odd sequences of digits (e.g., 7-8-3). Participants were instructed to press a specified key on the computer keyboard when they detected a signal. A total of 10 signals were presented in each 3-min period among a string of random digits for the high and low conditions (225 and 90 digits total, respectively). Workload level was

Downloaded from hfs.sagepub.com at UNIV OF OKLAHOMA on January 20, 2016

manipulated by changing the speed of digit presentation (i.e., high workload of 1 digit every 0.8 s vs. low workload of 1 digit every 2 s). In other words, the number of signals was the same for both workload conditions, but the total number of digits and the difficulty level varied according to the digit rate presentation. One question is whether signal probability and the number of digits to be processed per unit time in this study may be confounded. This potential limitation is more apparent than functional. Workload level was manipulated in terms of background event rate in this study and was consistent with procedures used in previous studies (e.g., see Warm & Jerison, 1984). Although, in general, performance typically varies inversely with event rate (Lanzetta, Dember, Warm, & Berch, 1987), there is evidence that increasing event rate does in fact result in increased perceived mental workload (Warm, Dember, & Hancock, 1996). Consequently, previous research on event rate manipulation would suggest that an increase in event rate would result in an increase in workload.

Procedure. When participants arrived at the laboratory, they were seated at individual workstations and asked to complete informed consent forms. Workstation partitions minimized distraction from other participants yet permitted individuals to view the experimenter during instruction administration.

Pilot study. Prior to this study, a pilot study ($N = 40$ participants) was conducted to determine the amount and type of training that would be needed to assure reasonably stable (or asymptotic) performance levels from most participants. The training regimen was patterned after that used by Schlegel and Gilliland (1990), which is relevant to visual and auditory tasks such the one employed in the present study. Schlegel and Gilliland (1990) conducted a comprehensive study of a variety of human performance tasks and identified training regimens that were sufficient to produce reasonably asymptotic performance. For example, they established that for visual and auditory tasks, thorough instructions followed by five to six 3-min trials resulted in stable performance. This pilot study confirmed that a thorough instruction set combined with six 3-min training trials (three 3-min trials at high workload and three 3-min at low workload, counterbalanced in their presentation) that included feedback to the participant was sufficient to attain reasonably steady performance levels. Thus, pilot data suggested that following the 18-min training session, participants understood and were performing the task well, therefore lessening the likelihood of test data contamination from learning effects.

Training and baseline data collection. Participants performed the Bakan Vigilance Task during three phases: training, baseline, and the experimental testing session. Table 1 presents the counterbalanced orderings for all training, baseline, and test trials. Each of the 198 participants was randomly assigned to one of the eight training-baseline-testing sequences in the high-low ($N = 104$) or low-high ($N = 94$) conditions. This resulted in not having exactly the same number of participants in each counterbalanced ordering.

Each participant was first familiarized with the location and operation of the computer response keys relevant for performing the task. Participants then engaged in the 18-min training session (three 3-min trials at high difficulty and three 3-min trials at low difficulty, counterbalanced in their presentation). Feedback was provided for the duration of the training protocol. Feedback

TABLE 1: Counterbalanced Orderings for Training, Baseline, and Test Trials for Experiments 1 ($N = 198$) and 2 ($N = 37$)

Training and Baseline Conditions								Test Condition (min)			
								Experiment 1		Experiment 2	
Training	Break	Training	Break	Baseline	Break	Baseline	Break	High to Low	Low to High	High	Low
Low, 9	5	High, 9	5	Low, 9	5	High, 9	15	9/9	9/9	18	18
High, 9	5	Low, 9	5	Low, 9	5	High, 9	15	9/9	9/9	18	18
Low, 9	5	High, 9	5	High, 9	5	Low, 9	15	9/9	9/9	18	18
High, 9	5	Low, 9	5	High, 9	5	Low, 9	15	9/9	9/9	18	18

Downloaded from hfs.sagepub.com at UNIV OF OKLAHOMA on January 20, 2016

for correct and mistaken signals was provided visually via the computer screen. A green pop-up box indicating "correct" or a red one indicating "incorrect" followed a participant's response. Furthermore, missed signals were indicated via a short auditory tone. The training regimen employed was again patterned after that used by Schlegel and Gilliland (1990). Training session trials were followed by an 18-min baseline session (three 3-min trials at high difficulty and three 3-min trials at low difficulty, in counterbalanced order) to establish baseline data for later comparisons. To minimize fatigue, 5-min rest breaks were given between each series of three training or baseline trials. During these breaks, participants were required to engage in a low-demand distractor task (i.e., completion of a demographic survey). The three trials of either training or baseline were each 3 min long and were presented with no discernable break between trials (i.e., appeared to be 9 continuous min). Following the second series of three baseline trials and prior to the testing session, participants were given a 15-min break, during which they completed the form mentioned previously.

Testing session. Following a 15-min break, during the test session, participants in the high-low condition engaged in three 3-min trials at high task difficulty followed immediately by three 3-min trials at low task difficulty. In contrast, participants in the low-high test condition engaged in three 3-min trials at low task difficulty followed immediately by three 3-min trials at high task difficulty. These two test conditions created a situation in which participants developed a workload history at one workload level and then moved immediately to a dramatically different workload level. The transitions between workload levels during the high-low and low-high test sessions were uninterrupted by rest periods (i.e., the total of six trials that included the shift in workload) and were thus perceived by the participant as 18 min of continuous work. Participants were run between the hours of 9:00 a.m. and 4:00 p.m. to control for time-of-day effects (Revelle, Humphreys, Simon, & Gilliland, 1980).

Results

For purposes of baseline and test data comparisons, median baseline scores were computed for each participant. In other words, from the appropriate set of baseline trials (low or high, given the participant's assignment to the high-low or low-high condition, respectively), the median trial (and/or second of three) was selected for comparison to performance on the three (3-min) subsequent test trials. In other words, after the three baseline scores were put in rank order, the second (or middle) score was considered the median baseline score.

Separate repeated measures ANOVAs were then conducted for each of the high-low and low-high test conditions (including the appropriate median baseline and the three test trials at high or low) for the dependent measures of correct responses, response time, and total errors, with alpha level controlled by Dunn's procedure. The violation of sphericity was addressed by specifying the models and using the Huynh-Feldt correction. Correct responses consisted of the number of correct responses following signal presentation. Response time was recorded only for correct responses. Total errors represented the combination of commission and omission errors. These repeated measures ANOVAs were significant for correct responses (CRs), response time (RT), and total errors (TEs) for the high-low condition – CR: $F(3, 309) = 5.41$, $p = .002$; RT: $F(2.76, 284.28) = 3.40$, $p = .021$; TE: $F(2.94, 302.82) = 8.19$, $p = .0001$ – and the low-high condition – CR: $F(2.76, 256.68) = 7.10$, $p = .0002$; RT: $F(2.46, 228.78) = 4.09$, $p = .009$; TE: $F(2.76, 256.68) = 6.08$, $p = .0008$.

Significant results were examined further using multiple contrasts. Specifically, median baseline scores were contrasted with each of the appropriate test trials. For example, in the sudden decrease condition (high-low), the median low-difficulty baseline score was compared with the three test trials at low difficulty. The mean correct responses, mean response time, and mean total errors data for each trial, as a function of condition, are presented in Figures 1 and 2.

To examine the immediate effects of workload history, a participant's performance on the appropriate median baseline trial was compared with that on the first test trial (Test Trial 1) following the sudden shift in workload. Examining the second and third trials provided a method for examining longer term or time course effects of workload history. Results of contrast analyses indicated significant differences between baseline and the first test trial for correct responses and total errors following a workload shift for the high-low condition – CR: $F(1, 103) = 11.05$,

Downloaded from hfs.sagepub.com at UNIV OF OKLAHOMA on January 20, 2016

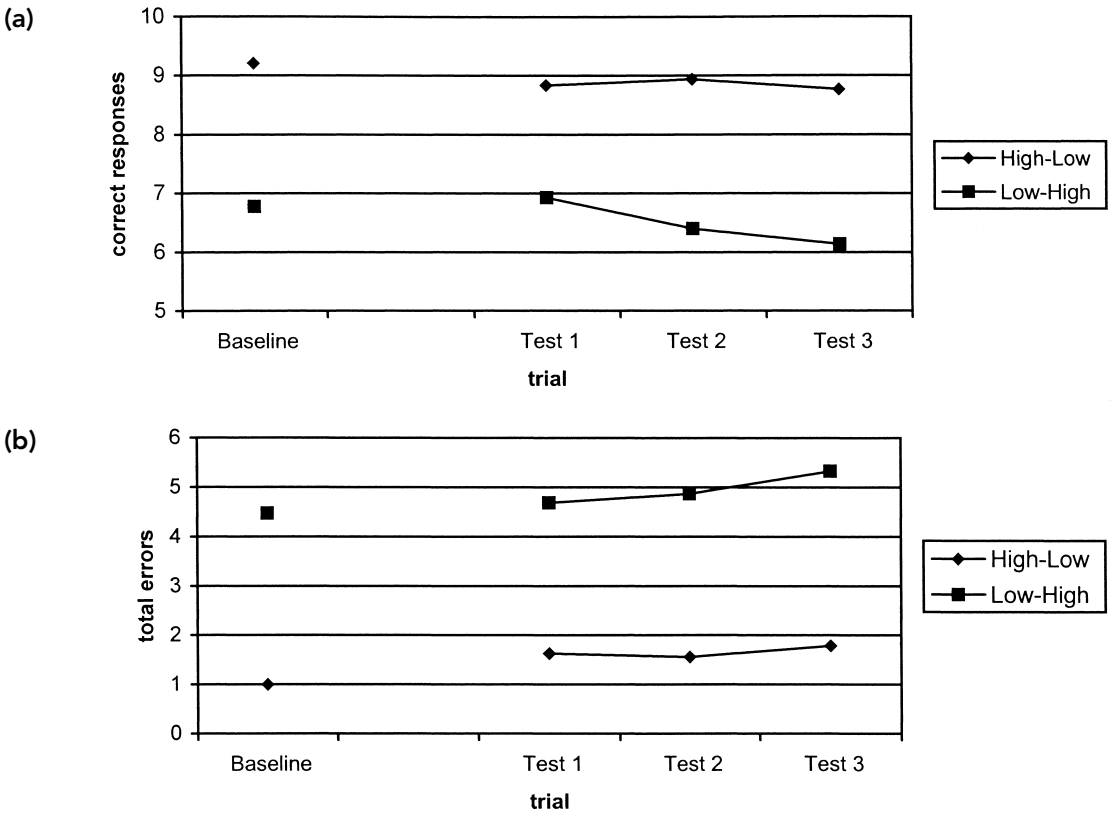


Figure 1. (a) Correct responses as a function of trial for the high-low ($N = 104$) and low-high ($N = 94$) conditions. For the high-low condition, Tests 1, 2, and 3 were significantly different from baseline data ($p < .017$). For the low-high condition, Test 3 was significantly different from baseline data ($p < .017$). (b) Total errors as a function of trial for the high-low and low-high conditions. For the high-low condition, Tests 1, 2, and 3 were significantly different from baseline data ($p < .017$). For the low-high condition, Test 3 was significantly different from baseline data ($p < .017$).

$p = .001$; TE: $F(1,103) = 15.71$, $p = .0001$ – but not for the sudden low-high condition – CR: $F(1, 93) = 0.48$, $p = .490$; TE: $F(1, 93) = 0.71$, $p = .402$. No significant differences were found for response time between baselines and the first test trial for either condition. Therefore, at least in terms of correct responses and total errors, there does appear to be a relatively immediate decrement in performance in the 3-min period following a sudden shift from high to low workload. In addition, it appears that an immediate shift from low to high workload did not result in an immediate statistically significant decrement in correct responses, total errors, or response time during this same initial 3-min period. Consequently, after correcting for the design and methodological limitations of previous studies, the results of this study provided evidence for the previously reported decrement in performance immediately following a transition from high to low workload.

The design of this study provided the opportunity to examine this decrement within a broader time course as well. Contrast analyses examined differences between baseline and the second and third 3-min trials (Trials 2 and 3) following the shift in workload. The correct response and total error data revealed the existence of a persistent significant decrement in performance following a shift in workload level for the high-low condition – CR: Trial 2 = $F(1, 103) = 6.24$, $p = .014$; CR: Trial 3 = $F(1, 103) = 12.50$, $p = .0006$; and TE: Trial 2 = $F(1, 103) = 8.51$, $p = .004$; TE: Trial 3 = $F(1, 103) = 20.22$, $p = .0001$. Contrast analyses examining differences between baseline and Trials 2 and 3 for the low-high condition revealed significant findings only for Trial 3 – CR: Trial 3 = $F(1, 93) = 14.50$, $p = .0003$; TE: Trial 3 = $F(1, 93) = 16.59$, $p = .0001$. The nature of this performance decrement varied somewhat in time course depending on the workload history. Although the

Downloaded from hfs.sagepub.com at UNIV OF OKLAHOMA on January 20, 2016

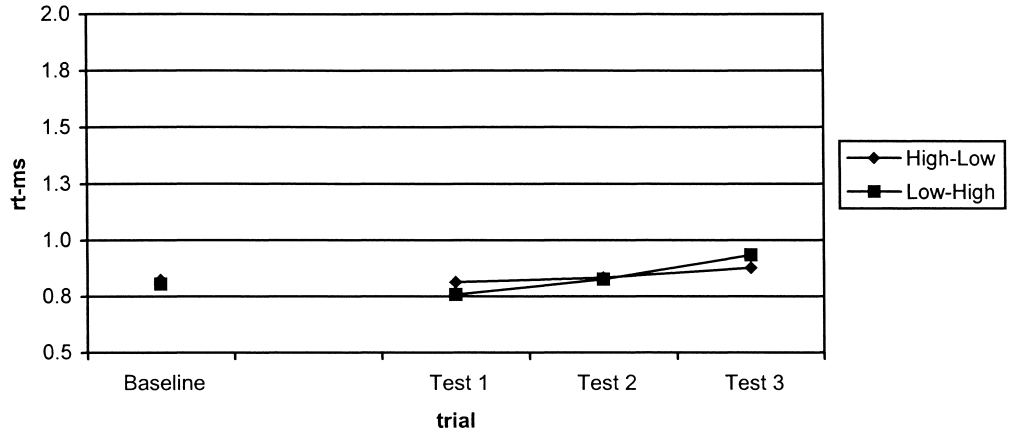


Figure 2. Response times as a function of trial for the high-low (N = 104) and low-high (N = 94) conditions. For both conditions, Test 3 was significantly different from baseline data ($p < .017$).

performance decrement in correct responses and total errors for the high-low condition was immediate, the performance decrement persisted for the full 9 min of test trials.

The performance decrement effect was also present, but delayed, for the low-high condition, emerging in the third testing trial. This is an important finding, given that previous research has suggested that a shift from a higher to a lower workload level results in a greater performance decrement as compared with a shift from a lower to a higher workload level (Cumming & Croft, 1973; Goldberg & Stewart, 1980; M. L. Matthews, 1986).

Response time was also investigated with contrast analyses. Although no immediate effects were found, analyses of differences between baseline and later test trials revealed decrements in RT performance on Trial 3 for both the high-low, $F(1, 103) = 8.61$, $p = .004$, and low-high test conditions, $F(1, 93) = 5.89$, $p = .017$. Immediate decrements in performance appear to occur after a reduction in workload, but either an increase or decrease in workload can lead to a loss in accuracy and a slowing of response time. It appears that for this task, however, the RT variable may be less sensitive than the CR and TE variables to the development of the decrement in performance.

Discussion

First, the findings of this study extend the time course of the performance decrement effect following a workload shift. Previous studies examining the effects of workload history limited their data sampling to much shorter periods of time following the workload change, ranging from 10 to 72 s (Cumming & Croft, 1973; Goldberg & Stewart, 1980; M. L. Matthews, 1986). This may explain the discrepancy in results between this and prior workload history studies. The present study suggested that the direction of the shift was less important than the shift itself, especially when considering longer term effects (i.e., both high-to-low and low-to-high workload shifts have longer term negative effects on performance). Thus, the dynamics of performance decrements associated with workload history are more complex than originally thought.

Furthermore, consider the two competing theoretical models offered to explain the decrement in performance after a high-to-low workload shift. M. L. Matthews (1986) suggested strategic persistence as an explanation for the decrement in performance. An alternative explanation suggested in this study is an adaptation-based model. It was noted that examining errors of omission or commission would provide a method to test which of these theoretical models best explains the decrement in performance. The absence of a response following the presentation of a signal resulted in an error of omission; errors of commission were recorded for responses in the absence of a signal. Errors of commission are more likely if a person is "overdriving" a task (as strategic persistence would suggest), whereas errors of omission are more likely if a person is seeking an opportunity to overcome a resource depletion state.

Downloaded from hfs.sagepub.com at UNIV OF OKLAHOMA on January 20, 2016

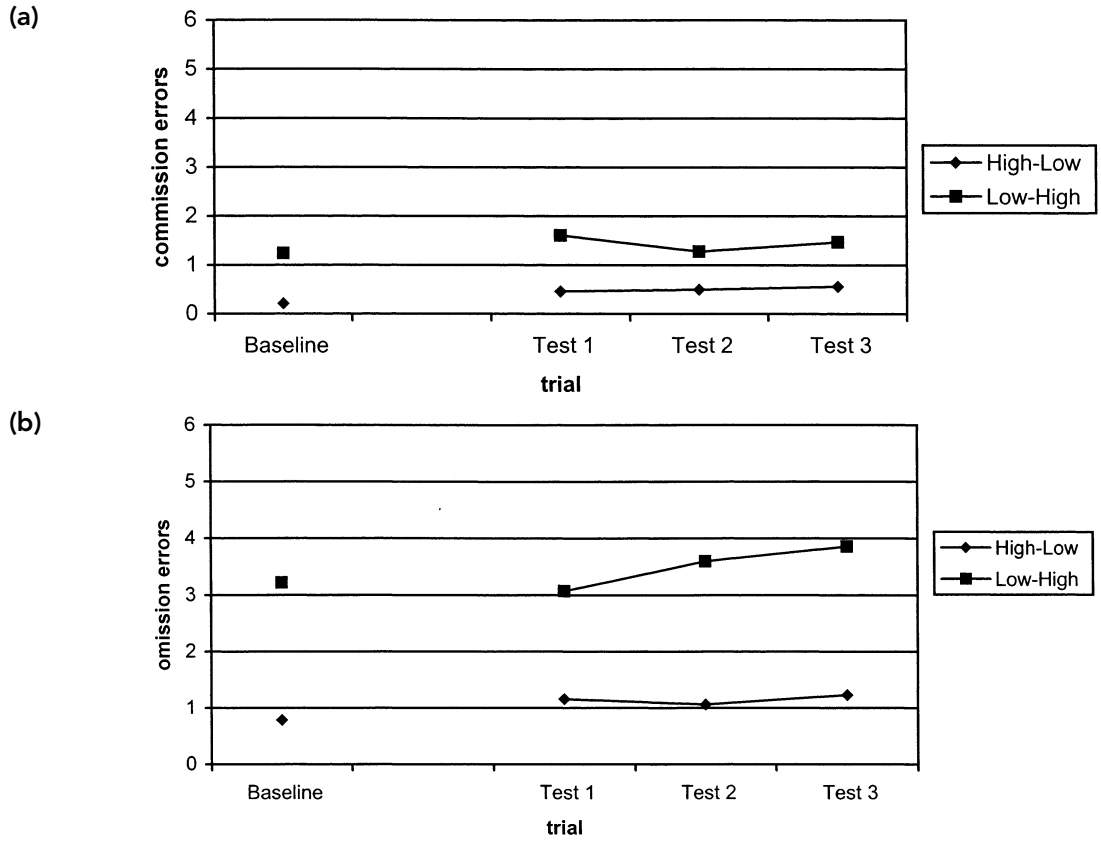


Figure 3. (a) Commission errors as a function of trial for the high-low ($N = 104$) and low-high ($N = 94$) conditions. For the high-low condition, Tests 1 and 3 were significantly different from baseline data ($p < .017$). For the low-high condition, test data were not significantly different from baseline data. (b) Omission errors as a function of trial for the high-low and low-high conditions. For the high-low condition, Tests 1, 2, and 3 were significantly different from baseline data ($p < .017$). For the low-high condition, Test 3 was significantly different from baseline data ($p < .017$).

To examine the nature of the performance decrement following a sudden decrease in workload level (high-low condition), repeated measures ANOVAs were conducted on errors of omission (OEs) and errors of commission (CEs) separately. The analyses identified statistically significant differences for both errors of omission, $F(3, 309) = 5.41, p = .002$, and errors of commission, $F(2.58, 265.74) = 4.34, p = .006$. The significant effects were investigated using multiple contrasts. Mean omission and commission error data for each trial are presented in Figure 3. The comparison of baseline scores and the first test trial provided the basis for investigating the nature of the immediate decrement following the sudden decrease in workload level – OE: Trial 1 = $F(1, 103) = 11.05, p = .0012$; CE: Trial 1 = $F(1, 103) = 6.60, p = .01$. In addition, in the high-low condition, contrast analyses between appropriate baselines and test trials were also found to be significant for errors of commission in Trial 3 – CE: Trial 3 = $F(1, 103) = 12.40, p = .0006$ – and errors of omission in Trials 2 and 3 – OE: Trial 2 = $F(1, 103) = 6.24, p = .014$; OE: Trial 3 = $F(1, 103) = 12.50, p = .0006$.

These contrasts revealed significant negative effects for both errors of omission and commission. Given the nature of the findings, it is difficult to unambiguously discriminate between the two theories. However, it seems that a significant increase in omissions is indeed incompatible with an explanation based on strategic persistence. If, in fact, participants were maintaining unnecessarily high levels of effort, it is unlikely that they would commit significantly more errors of omission. More importantly, a significant increase in errors of commission might not be incompatible with a theoretical explanation based on

Downloaded from hfs.sagepub.com at UNIV OF OKLAHOMA on January 20, 2016

resource adaptation. In fact, from a resource adaptation perspective, if participants were responding more slowly to signals, these late correct responses would have been incorrectly registered as errors of commission when in fact they were merely slow, perhaps because of resource depletion. Unfortunately, the constraints on the data collection method do not permit a definitive test of this post hoc hypothesis; however, the data suggest that a theoretical explanation based on resource adaptation is more likely.

Furthermore, the resource adaptation perspective offers another explanation for the results of this study – one based on motivation. Specifically, participants might have demonstrated a simple lack of motivation attributable to a state of depleted resources. The diversion of attention (or distraction) might be responsible for the significant increase in both commission and omission errors. In other words, it is possible that loss of motivation might result in random or reduced key pressing producing significantly more errors of both types. It is also possible that boredom may have played a role in the performance declines. Specifically, Scerbo (1998) suggested that boredom produced by a repetitive task over which participants have no control results in performance decline.

The Questions of Fatigue or Boredom

Based on the results of this study, a logical question would be the extent to which time on task or fatigue might have influenced the results. This is an important question given the abundance of research suggesting that sustained attention on some tasks may result in a significant vigilance decrement (e.g., Davies & Parasuraman, 1982; Mackworth, 1948, 1950/1961; Warm, 1984). However, it is important to note that some research suggests that the vigilance decrement is not as likely with cognitive tasks (such as the Bakan Vigilance Task) as it is with sensory tasks (See, Howe, Warm, & Dember, 1995).

Furthermore, given the findings of Scerbo (1998), another logical question might be the extent to which boredom may have affected the findings. A follow-up study was conducted to investigate the presence of a time-on-task or fatigue effect, as well as a potential boredom effect. It was expected that if fatigue or boredom were a significant factors, the effects might be present (if not more pronounced) over a prolonged period at a sustained level of workload. Consequently, the workload history design used required participants to work at a fixed workload level following a training-baseline protocol identical to that of the original study (see the right two columns of Table 1).

FOLLOW-UP STUDY

Method

Participants. Participants were 40 individuals randomly selected and recruited from approximately 450 available volunteer participants. Participants for this study were contacted by telephone and invited to participate. Of those contacted, 3 participants failed to complete the experiment (1 was dropped for failing to understand the task and 2 were dropped because of equipment failure), leaving 37 undergraduate students (18 men and 19 women) from the University of Oklahoma. Participants received extra credit as one option for fulfilling the requirements of undergraduate lower division psychology courses.

Procedure. The methods for the second study are essentially the same as those in the first study (i.e., materials, training, and baseline data collection), with the exception of the testing session. In Study 2, a participant's assignment to the high (N = 18) or low (N = 19) testing condition was based on the training-baseline-testing protocol sequence to which he or she was originally randomly assigned (see Table 1). During the test session, participants in the high condition engaged in six 3-min trials at high task difficulty. In contrast, participants in the low test condition engaged in six 3-min trials at low task difficulty. These two test conditions created a situation in which participants developed a workload history at a single workload level. Transitions between 3-min trials were uninterrupted by rest periods. Thus, the transition times between 3-min trials for the high and low test conditions were no different than those within the series of high or low trials (i.e., in the high condition the six 3-min test trials at high workload was perceived by the participant as 18 min of continuous work).

Results

The median baseline scores and test trial data were used in this analysis much as in Study 1. Separate repeated measures ANOVAs were conducted

Downloaded from hfs.sagepub.com at UNIV OF OKLAHOMA on January 20, 2016

for correct responses, response time (for correct responses), and total errors measures for the continuous high and low workload conditions. These analyses yielded no significant time-on-task effects for correct responses, response time, or total errors for either the high condition – CR: $F(6, 102) = 0.97$, $p = .484$; RT: $F(6, 102) = 1.85$, $p = .172$; TE: $F(6, 102) = 1.47$, $p = .267$ – or the low condition – CR: $F(4.86, 87.48) = 2.20$, $p = .110$; RT: $F(4.98, 89.64) = 1.02$, $p = .461$; TE: $F(5.76, 103.68) = 0.796$, $p = .589$. The correct responses, response time, and total errors data for each trial are indicated in Figures 4 and 5 (see Table 2 for high condition effect sizes). Therefore, neither fatigue nor boredom was considered a significant factor in this experiment, and both are considered unlikely to have been significant factors in Study 1.

GENERAL DISCUSSION

The purpose of this project was threefold. First, this study examined the effects of workload history while addressing several design and methodological limitations that posed threats to the internal and external validity of previous studies. The results indicated, at least in terms of correct responses, a relatively immediate decrement in performance followed a sudden shift from high to low workload, but no immediate decrement followed a sudden shift from low to high workload. However, this study suggests that both a sudden decrease and a sudden increase result in negative effects over a longer time course.

This study was also designed to test two competing theoretical approaches (one based on strategic persistence, the other based on adaptation) proposed to explain the nature of responses to these changes in workload level. The findings made it difficult to discriminate definitively between the two approaches, but post hoc interpretations for both immediate and longer term effects favored an adaptation theoretical perspective.

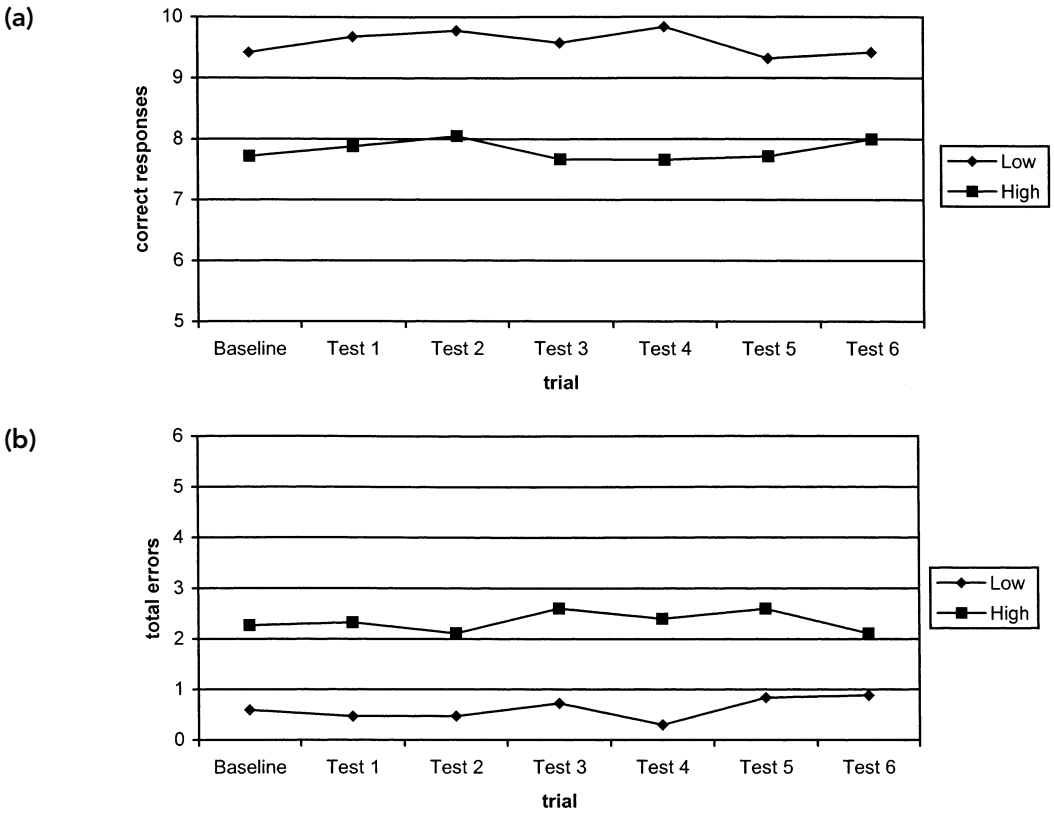


Figure 4. (a) Correct responses as a function of trial for the high ($N = 18$) and low ($N = 19$) conditions. For both conditions, test data were not significantly different from baseline data. (b) Total errors as a function of trial for the high and low conditions. For both conditions, test data were not significantly different from baseline data.

Downloaded from hfs.sagepub.com at UNIV OF OKLAHOMA on January 20, 2016

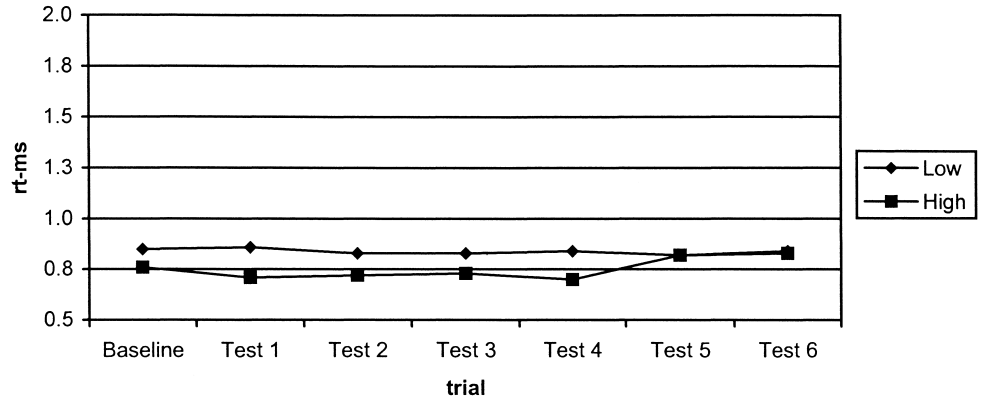


Figure 5. Response times as a function of trial for the high ($N = 18$) and low ($N = 19$) conditions. For both conditions, test data were not significantly different from baseline data.

Previous workload history theories do not provide predictive efficiency, regardless of the direction of the workload shift; however, an alternative explanation centering on an adaptation-based theory may have broader predictive or explanatory value. Predictions for the high-low condition based on such a theory have already been noted. Interestingly, such a theory could also address the result found in the low-high condition. This explanation implies that a modest period of low workload would not seriously deplete resources. The initial increase in performance efficiency following a shift to a higher workload level, as was seen in the low-high condition (see Figure 3b, Trial 1), might be a result of recruiting the necessary resources to meet the demand of the task. Although resources are being used at a low workload level, significant depletion is unlikely. Consequently, when a shift occurs, the individual is initially able to recruit the resources necessary to meet the demands of the task and perform at the same or higher level. An adaptation-based theory (e.g., the state control theory of adaptation; Hockey, 1986, 1997) would also suggest that as the high workload level continued, the later decrement in performance (Trial 3) was a result of resource depletion and/or subsequent recuperative efforts. Thus, following the longer period at the high workload level, increased resource depletion became evident. In contrast, the strategic persistence model fails to provide a satisfactory explanation for the initial improvement in performance followed by the gradual decrement in the omission error data for the low-high condition. In fact, the strategic persistence model predicts significantly more (not fewer) omissions following the immediate shift from the low to high workload level.

One logical question would be the degree to which fatigue or boredom could have influenced the findings of this study. However, the follow-up study investigated the extent to which these factors might have been responsible for the decrement in performance following sudden transitions in

TABLE 2: Effect Sizes for Response Time in High Condition

	M (ms)	*SD* (sd)	Effect Size (*r*)	Cohen's *d*
Baseline	0.76	0.15		
Test 1	0.71	0.18	.1492	.3017
Test 2	0.72	0.18	.1198	.2414
Test 3	0.73	0.17	.0931	.1871
Test 4	0.7	0.13	.2090	.4274
Test 5	0.82	0.21	.1622	.3287
Test 6	0.83	0.29	.1498	.3032

Downloaded from hfs.sagepub.com at UNIV OF OKLAHOMA on January 20, 2016

workload (reported in the original study), and it yielded no significant findings for participants working at either high or low workload level over a prolonged period. Thus, it appears that the decrement in performance following workload transitions might be a consequence of something inherent in the shift rather than an effect of fatigue or boredom.

It is becoming increasingly apparent that a simple explanation addressing responses to a change in workload level fails to address the complexities involved in the more dynamic effect of workload history on performance. The importance of this line of research is evident for many real-world jobs, particularly those involving safety-sensitive occupations. The immediate response of a bank security guard suddenly confronted with a holdup, or of an emergency room doctor with significantly fewer patients to attend, can have life-or-death implications for those involved. These scenarios involve significant workload transitions similar in a general sense to those presented in this study. Consequently, improvements in methodology (i.e., adjustments in task difficulty) and changes in types of tasks need to be considered in future studies if researchers are to come to a better understanding of the effect of workload history on performance. In addition, further exploration of factors potentially responsible for postshift detrimental effects need to be further investigated. For example, depending on the task, it may be helpful to examine whether participants, following sudden increases and decreases, engage in inappropriate observing strategies at the new workload level. Future studies may want to test the predictive efficiency of potentially competing explanatory theoretical models, such as a resource theory view and an effort regulation model (see G. Matthews & Desmond, 2002). Furthermore, the relationship between individual differences or personality and workload history might have considerable promise for expanding understanding of the dynamic influences of workload history on performance.

Finally, the complexities involved in the dynamic effects of workload history on performance, along with the many workable workload history scenarios, necessitate a reconceptualization of "workload history." Indeed, it appears that workload history may provide a basic structural framework for an expanded exploration of human performance research.

ACKNOWLEDGMENTS

The author expresses her gratitude to Drs. Kirby Gilliland, Larry Toothaker, and Scott Gronlund for their many helpful suggestions and to the members of the Personality and Performance Research Lab for their help as research assistants.

REFERENCES

Bakan, P. (1959). Extraversion-introversion and improvement in an auditory vigilance task. *British Journal of Psychology, 50,* 325–332.

Cumming, R. W., & Croft, P. G. (1973). Human information processing under varying task demand. *Ergonomics, 16,* 581–586.

Davies, D. R., & Parasuraman, R. (1982). *The psychology of vigilance.* London: Academic Press.

Goldberg, D. R., & Stewart, M. R. (1980). Memory overload or expectancy effect? "Hysteresis" revisited. *Ergonomics, 23,* 1173–1178.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31,* 519–537.

Hockey, G. R. J. (1986). A state control theory of adaptation and individual differences in stress management. In G. R. J. Hockey, A. W. K. Gaillard, & M. G. H. Coles (Eds.), *Energetics and human information processing* (pp. 285–298). Dordrecht, Netherlands: Kluwer Academic.

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology, 45,* 73–93.

Huey, B. M., & Wickens, C. D. (Eds.). (1993). *Workload transition: Implications for individual and team performance.* Washington, DC: National Academy Press.

Hunt, R. R., & Ellis, H. C. (2004). *Fundamentals of cognitive psychology.* New York: McGraw-Hill.

Krulewitz, J. W., Warm, J. S., & Wohl, T. H. (1975). Effects of shifts in the rate of repetitive stimulation on sustained attention. *Perception and Psychophysics, 18,* 245–249.

Lanzetta, M., Dember, W. N., Warm, J. S., & Berch, D. B. (1987). Effects of task type and stimulus heterogeneity on the event rate function in sustained attention. *Human Factors, 29,* 625–633.

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology, 1,* 6–21.

Mackworth, N. H. (1961). Researches on the measurement of human performance. In H. W. Sinaiko (Ed.), *Selected papers on human factors in the design and use of control systems* (pp. 171–331). New York: Dover. (Original work published 1950)

Matthews, G., & Desmond, P. (2002). Task-induced fatigue states and simulated driving performance. *Quarterly Journal of Experimental Psychology, 55,* 659–686.

Matthews, M. L. (1986). The influence of visual workload history on visual performance. *Human Factors, 28,* 623–632.

Moroney, B. W., Warm, J. S., & Dember, W. N. (1995). Effects of demand transitions on vigilance performance and perceived workload. In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (pp. 1375–1379). Santa Monica, CA: Human Factors and Ergonomics Society.

Revelle, W., Humphreys, M., Simon, L., & Gilliland, K. (1980). The interactive effect of personality, time of day, and caffeine: A test of the arousal model. *Journal of Experimental Psychology: General, 109,* 1–26.

Scerbo, M. W. (1998). What's so boring about vigilance? In R. R. Hoffman, M. F. Sherrick, & J. S. Warm (Eds.), *Viewing psychology as a whole: The integrative science of William Dember* (pp. 145–166). Washington, DC: American Psychological Association.

Schlegel, R. E., & Gilliland, K. (1990). *Evaluation of the Criterion Task Set – Part I CTS performance and SWAT data-baseline conditions (U)* (Tech. Rep. AAMRL-TR-90-007). Wright-Patterson Air Force Base, OH: Aeromedical Research Lab.

Downloaded from hfs.sagepub.com at UNIV OF OKLAHOMA on January 20, 2016

See, J., Howe, S., Warm, J., & Dember, W. (1995). Meta-analysis of the sensitivity decrement in vigilance. *Psychological Bulletin, 117,* 230–249.

Selye, H. (1956). *The stress of life.* New York: McGraw-Hill.

Selye, H. (1978). *The stress of life* (Rev. ed.). New York: McGraw-Hill.

Warm, J. S. (1984). An introduction to vigilance. In J. S. Warm (Ed.), *Sustained attention in human performance* (pp. 1–14). Chichester, UK: Wiley.

Warm, J. S., Dember, W. N., & Hancock, P. A. (1996). Vigilance and workload in automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 183–200). Mahwah, NJ: Erlbaum.

Warm, J. S., & Jerison, H. J. (1984). The psychophysics of vigilance. In J. S. Warm (Ed.), *Sustained attention in human performance* (pp. 15–59). Chichester, UK: Wiley.

Luz-Eugenia Cox-Fuenzalida is an assistant professor in the Department of Psychology at the University of Oklahoma, where she received her Ph.D. in psychology in 2000.

Date received: January 9, 2003
Date accepted: December 23, 2005

Downloaded from hfs.sagepub.com at UNIV OF OKLAHOMA on January 20, 2016