

## **Effect of Training on Workload in Flight Simulation Task Performance**

**Indramani L. Singh, Hari Om Sharma and Anju L. Singh**  
Banaras Hindu University, Varanasi

The present study examines the effect of training on workload in flight simulation task performance. A 2(training) x 2(session) x 3(block) mixed factorial design was used. Training was treated as between subjects factor, whereas sessions and blocks were treated as within subject factors. A revised version of multi-attribute task battery was used in this study. Thirty under graduate students of the Banaras Hindu University served as subjects. Each subject was required to detect automation malfunctions within stipulated time i.e. 10 sec. and to reset errors, if any, by pressing a designated key. The performance was recorded in terms of accuracy of target detection (hit rates), incorrect detection (false alarms) and reaction time. The NASA-TLX (Hart & Staveland, 1988) was used for the assessment of workload having six bipolar dimensions like mental demands, physical demands, temporal demands, effort, frustration and own performance before and after a 30-min (short) and a 60-min (long) sessions of manual training. Mean detection performance showed higher hit rates in long training session than in short training session. However, the mean difference between two training conditions was not found to be significant. Thus results suggested that the amount of training did not affect subjects' system-monitoring task performance under automated mode. The main effects of session showed significant decrement in the detection of automation failures over sessions. Results also indicated that subjects reported significantly higher temporal workload between pre-and post test session in short training than long training condition. Further subjects showed significantly high degree of frustration workload in pre than post automated task performance in long training condition

Automation has encapsulated almost every domain of human life. Many automated systems have replaced the human component as a key element. Moreover, it is well established that automation can perform functions more efficiently, reliably, or accurately than the human operator. Monitoring of highly automated systems is a major concern for human performance efficiency and system safety in a wide variety of human-machine system. Human operators using modern human-machine systems are faced with increased demands on

monitoring and supervisory control because of greater levels of automation (Parasuraman, 1987; Wiener, 1988). In general, human operators perform well in the diverse working environments, especially in the area of air traffic control, surveillance operations, power plants, intensive care units, and quality control in manufacturing. But it does not mean that errors do not occur. A study conducted by the US Air Transport Association on personnel, who conducted x-ray screening for weapons at airport security checkpoints recorded their monitoring

performance. The detection rate of these weapons and explosives was typically good but not perfect. There could be several reasons for imperfect performance like unhealthy working environment (noisy and distracting), uncomfortable sitting arrangement in adverse visual conditions, lack of training, low wages, or lack of opportunities for career advancement.

These technological advances have also been associated with certain behavioural problems like a loss of situation awareness, deregulation of mental workload (Parasuraman, 1987). A potential cost of automation has also been noted like 'automation-induced complacency' (Parasuraman, Molloy, & Singh, 1993).

Parasuraman et al. (1993) empirically validated the concept of automation-induced complacency. They found automated complacency in the multi-task environment rather than in the single task situation. In a follow-up study, Singh, Parasuraman, Deaton, & Molloy, (1993) also reported automated complacency on U.S. pilots. Furthermore, Singh, Molloy and Parasuraman (1993) developed 'complacency potential rating scale' (CPRS) which had trust, confidence, and reliance as main factors. Accordingly they suggested that pilots 'trust, confidence and reliance in system might affect his or her use of automation.

Moreover, automation has been designed with the objective to reduce operator's mental workload; however, results suggested that automation does not necessarily reduce workload (Parasuraman, & Riley, 1997; Singh, & Parasuraman, 2001). Parasuraman (1999) suggested that automation could reduce human operator's workload to an optimal level, if it is suitably designed. Further, if automation is implemented in a 'clumsy manner', workload reduction may not occur (Wiener, 1988). The first evidence appeared from a survey of commercial pilots and their attitudes toward advanced cockpit automation (Madigan, &

Tsang; 1990; Wiener, 1988) that workload may not be necessarily reduced. It could be because, firstly, automation may change the pattern of workload across work segments. Second, the demands of monitoring can be considerable (Parasuraman, Mouloua, Molloy, & Hilburn, 1996).

Workload can be characterized as the interaction between the components of the machine and task on the one hand and the operator's resource capabilities, motivation and the state of the mind on the other hand (Gopher & Donchin, 1986; Moray, 1988; Wickens, & Kramer, 1985). A more precise definition of workload has been defined as the 'costs' a human operator incurs to complete an assigned task (Kramer, 1991). Automation, if properly designed, can reduce human operator's workload to a manageable level under peak loading conditions. Braby, Harris, and Muir (1993) reported that high levels of workload could lead to errors and system failures, whereas low workload could lead to complacency. It could be a reason for using automation in the first place to reduce high demands on the operator, resulting decrement in human errors. The findings further demonstrated that high automation often redistributed rather than reduced the workload within the system (Lee & Moray, 1992; Parasuraman & Mouloua, 1996; Singh et al., 2001; Wiener, 1988). This is because automation sometimes poses high demand and that become difficult to cope or manage, and which may require more attention.

Some of the studies have examined the effects of automation and workload on monitoring/tracking performance. Wickens, and Kessel, (1981) studied the effect of workload on active and passive control of tracking and monitoring system. Results showed significantly poor detection of failures in the automated than in the manual condition. Similar results were obtained by

Johannsen, Pfender, and Stein, (1976), wherein subjects were required to detect autopilot failures during a landing approach with simulated vertical gusts and in a study where pilots had to respond to unexpected wind-shear bursts while using an automated latitude hold. Contrarily, Liu, Fuld, and Wickens (1993) found that error detection was superior in the automated than in the manual condition. Hilburn, Jorna, and Parasuraman (1995) also found benefits of automation use on monitoring performance.

However, none of the researcher so far attempted to examine the effect of extended manual training on workload and automated monitoring performance. Some of the studies reported association or dissociation for brief performance periods after small amount of practice. Thus several practical consideration emanate from this expectation for future researches in the area of man-machine interaction. In the present study an attempt has been made to examine whether automation reduce workload in multi-task environment with increased amount of manual training.

### **Hypotheses**

The following hypotheses have been framed for testing in this study:

- 1) Amount of training would reduce automated complacency.
- 2) Increased training would reduce mental workload
- 3) Automated complacency would be progressively higher across time periods.

### **Method**

Thirty graduate students of the Banaras Hindu University were randomly employed in this study. Each subject had normal (20/20) or corrected to normal visual acuity, and their age varied from 19 to 22 years. None of the subject had prior experience of flight simulation task.

### **Flight Simulation Task**

A revised version of Multi-Attribute Task Battery (MAT: Comstock & Arnegard, 1992) was used in the present study which comprised system-monitoring, tracking, and fuel resource management. Out of these three tasks, only system-monitoring task was automated and remaining two tasks i.e. tracking, and fuel-management tasks were performed manually. These tasks were displayed in separate windows on a 14" SVGA colour monitor of a computer.

### **System-Monitoring Task**

The system-monitoring task, having four vertical gauges with moving pointers and green ok and red warning lights, was presented in upper left window of the monitor. The scales for gauges were marked for temperature and pressure of two engines of an aircraft. In normal condition, the green ok light remains on and the pointers fluctuate .25" from the centre of the gauges. Occasionally, a system malfunction occurred on one of the four gauges and the pointer went off the limits. Under automation mode, these malfunctions were detected and corrected automatically by system. However, from time to time, the automation failed to detect the malfunctions, and the subject was required to reset these malfunctions by pressing one of the corresponding function keys on the keyboard. All participants were required to detect malfunctions within 10 sec, failing which score would be recorded as miss. If the subject responded correctly within required time, score was recorded as hit score. Thus the performance measures were the correct detection (hit rates), incorrect detection (false alarms) and reaction time (RT).

### **Tracking Compensatory Task**

A two dimensional compensatory tracking task with joystick control was presented in central window of the monitor. A circular target symbol representing the deviation of aircraft

from its course fluctuated within the window. The subject's task was to keep the target symbol within the central rectangular area by applying the appropriate joystick's control inputs. Root mean square (RMS) error was computed as tracking performance score.

### **Fuel Resource Management Task**

This task comprised six fuel tanks duly linked with eight pumps. The amount, rate and direction of fuel flow were displayed on the monitor. The subjects were instructed to maintain the fuel level at 2500 gallons, each in tank A and tank B, which consumed the fuel at the rate of 800 gallons per minute. The subjects could take the required fuel from lower supply tanks by activating the appropriate pump. All pumps had different pumping capacity and these could be turned on or off by pressing the corresponding numbered keys from 1 to 8 on the keyboard. A root mean square error (RMS) was computed as performance measure in this task.

### **Assessment of Workload**

The NASA task load index (NASA-TLX; Hart, & Staveland, 1988) was used for the assessment of workload. The NASA-TLX is considered to be one of the most effective measures of perceived workload so far (Nygren, 1991). This scale has very high reliability ( $r = .83$ ) and it has six sources of workload. Three of those sources reflect the demands which experimental tasks place on operator (mental, physical, and temporal demand), whereas remaining three sources characterize the interaction between the operator and the task (effort, frustration, and performance). This scale provides an overall workload score based on a weighted average of ratings.

### **Design**

A 2(training) x 2(session) x 3(block) mixed factorial design was used in the present study.

Training was treated as between subjects factors, whereas sessions and blocks were treated as within subjects factors. The subjects were assigned randomly to either short or long training groups under static reliability (87.5%) condition from block to block.

### **Procedure**

All subjects were required to fill-up a consent form individually for participating in this experiment. They were also required to complete a biographical questionnaire, which had several questions about their age, education, socio-economic status, knowledge about computer and frequency of practice on a computer. Subjects were also tested for their normal vision on Snellon vision chart in the lab. All thirty subjects received a 10-min common practice on all three subtasks, i.e., system-monitoring, tracking and fuel management of flight simulation task. At the end of practice, feedback was given to each participant on dependent measures. Thirty subjects, who scored about 60% or above on system-monitoring task in practice session, were selected for final test sessions. Out of 30 subjects, 15 subjects were given short manual training (30-min) and 15 subjects were given long manual training (60-min) on flight simulation task. Feedback of performance for all the three tasks was given to the subjects. After short/long training, all subjects were required to perform final test session, comprising two 30-min sessions.

NASA-Task Load Index was administered individually to all subjects once after training session and once after the automated task was completed.

### **Results and Discussion**

The means and standard deviations for hits rate, false alarms, reaction time, and root mean square errors response criteria were computed separately for the common 10-min practice, short manual and long manual

training and final six 10-min automated test sessions.

### Practice Performance

Results indicated that the mean correct detection of malfunctions varied from 64% to 67% between short and long manual training conditions. However, mean difference in hits rate performance was not significant between short and long training. Similar results were obtained for F.As. and R. T. performance measures. The tracking, and fuel resource management RMS errors performance also did not show significant difference from short to long training conditions. Results indicated the equivalence of initial performance levels between two training conditions.

### Training Performance

The mean difference on all five dependent measures after a 30-min or 60-min manual training revealed that the mean

correct detection (hit rates) of short training and long training varied from 65% to 70%, respectively. However, the difference was not significant. This trend of results was maintained for the remaining other measures like false alarms, reaction time and RMS performances. Thus, results suggested that the extended training did not improve performance.

### Automation Task Performance

Mean correct detection performance showed a marginally higher in long training condition ( $M = 41.67$ ,  $SD = 5.51$ ) than short training ( $M = 37.82$ ,  $SD = 5.51$ ). However, this difference was not found significant. Data were then subjected to a 2 (training) x 2 (session) x 3 (block) analysis of variance with repeated measures on the last two factors, summary of which is presented in Table 1.

**Table 1: Summary of Analysis of Variance for Correct Detection of Malfunctions in Short and Long Training Conditions.**

| Source                 | Sum of Squares | df | Mean Squares | F-values |
|------------------------|----------------|----|--------------|----------|
| Training               | 665.08         | 1  | 665.08       | 0.243    |
| Error                  | 76767.49       | 28 | 2741.69      | —        |
| Session                | 6432.08        | 1  | 6432.08      | 4.96*    |
| Session*Training       | 3.2            | 1  | 3.2          | 0.002    |
| Errors                 | 36272.38       | 28 | 1295.44      | —        |
| Block                  | 4886.04        | 2  | 2443.02      | 2.13     |
| Block*Training         | 2099.37        | 2  | 1049.68      | 0.91     |
| Error                  | 64209.24       | 56 | 1146.59      | —        |
| Session*Block          | 387.37         | 2  | 193.68       | 0.221    |
| Session*Block*Training | 422.93         | 2  | 211.46       | 0.241    |
| Error                  | 49051.02       | 56 | 875.91       | —        |

The ANOVA results showed that main effect of training was not found significant,  $F(1, 28) = 0.24$ ;  $p > .62$ , which revealed that amount of manual training given prior to the detection of automation failures has no impact on 'automated complacency'. The

interactions were also not found significant. In this experiment, only 15 subjects were assigned in each training group, so the size of sample could be a factor for obtaining insignificant main effect of training. Thus, the result does not support the first



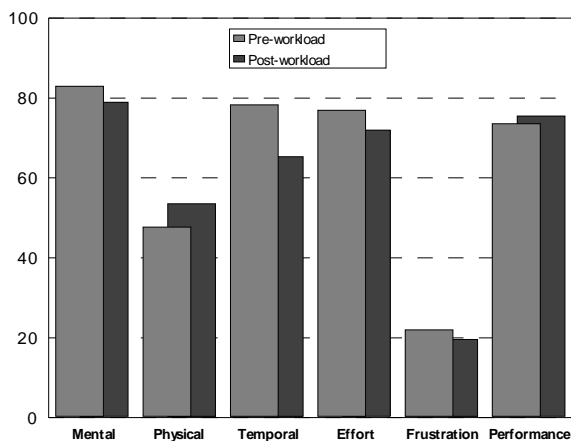
hypothesis that the amount of training would reduce automated complacency.

The significant main effect of session,  $F(1, 28) = 4.96$ ;  $p < .05$  showed significant decrement in detection of automation failures from session to session. This result supported the second hypothesis that detection of automation failures would progressively decline over sessions. The findings are consistent with other researchers (Parasuraman et al. 1993; Singh et al, 1997) who also reported automated complacency over sessions under multi-tasks environment.

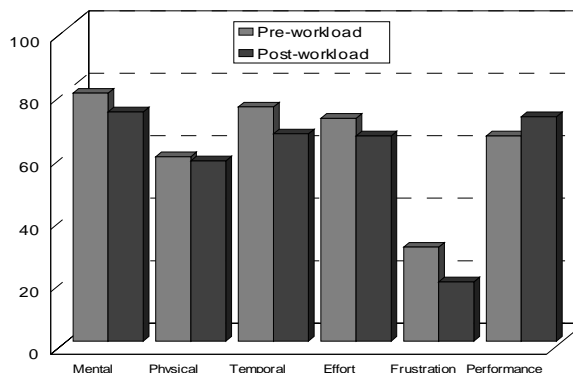
**Automation and Perceived Workload**

The mean scores of the perceived workload dimensions indicated that pre-mental, pre-temporal, pre-effort and pre-frustration indices were high in both short and long training condition than their counterpart conditions i.e. post-mental, post-temporal, post-effort and post-frustration. The obtained mean scores are graphically presented in Figure 1 and 2 for short and long training conditions.

**Figure 1. Pre and Post Mean Workload Scores in Short Training Condition**



**Figure: 2. Pre and Post Mean Workload Scores in Long Training Condition**



These mean differences were further compared by using paired compared t-test, since these workload scores were obtained at two intervals of time i.e. before and after automated monitoring task performance in both short and long training groups. The obtained t-values for short and long training are presented separately in Table 2.

**Table 2. Summary of the Paired Sample t-test Values Between Pre- and Post Test Sessions after Short and Long Training Conditions (N = 15).**

| Workload dimension                           | Short training t-value | Long training t-value |
|--|------------------------|-----------------------|
| Pre-Mental Demand vs Post-Mental Demand      | 1.20                   | 1.59                  |
| Pre-Physical Demand vs Post-Physical Demand  | 1.26                   | 1.17                  |
| Pre-Temporal Demand vs Post -Temporal Demand | 3.23**                 | 1.22                  |
| Pre-Effort vs Post -Effort                   | .01                    | 1.3                   |
| Pre-Frustration vs Post -Frustration         | .72                    | 2.07*                 |
| Pre-Own Performance vs Post-own Performance  | 0.70                   | 1.08                  |

\*\* p<0.01 \*p<0.01

Results suggested that the amount of workload of all dimensions decreased in post test session after performing automated monitoring task but significant difference emerged only in temporal workload in short training and frustration workload in long training group. Paired t-test was also computed on the mean difference in pre-and post workload of all 30 participants, irrespective of the amount of manual training. These findings (see Table 3) demonstrated that automation significantly reduced mental ( $t = 2.02$ ;  $p < .05$ ), temporal ( $t = 2.70$ ;  $p < .01$ ) and frustration workload ( $t = 2.11$ ;  $p < .04$ ). However, the other dimensions of workload i.e. physical, effort and performance did not show significant difference which could be due to small sample size i.e. 30 subjects.

**Table 3. Paired Sample t-test on the Total Sample (N=30)**

| Workload dimension                               | t-value | df |
|--|---------|----|
| Pre-Mental Demand vs<br>Post-Mental Demand       | 2.02*   | 29 |
| Pre-Physical Demand vs<br>Post-Physical Demand   | 0.50    | 29 |
| Pre-Temporal Demand vs<br>Post - Temporal Demand | 2.70**  | 29 |
| Pre-Effort vs Post -Effort                       | 0.927   | 29 |
| Pre-Frustration vs<br>Post -Frustration          | 2.11*   | 29 |
| Pre-Own Performance vs<br>Post-own Performance   | 1.29    | 29 |

\*\*  $p < 0.01$  \* $p < 0.05$

The obtained results partially support our assumption that automation would reduce mental workload. The high reliability of automated system use in the modern cockpit may reduce some of the workload of the pilot, but it may actually increase other components of workload (Masaloni, Duley

& Parasuraman, 1999). However, this trend of result may not be true in all the case. There is behavioural and subjective evidence for automation increasing workload, probably due to the operator having to monitor whether the automation is functioning correctly (Parasuraman & Riley, 1997; Wickens, Mavor, & McGee, 1977).

Thus the results of the present study are relevant to the debate on technology-centered versus human-centered approaches to the design of automation. The dominant tendency of the technology-centered approach has been to implement automation wherever possible in order to reduce pilot workload and to reap the benefits of economies such as fuel efficiency and reduced training cost.

## References

- Braby, C. D., Harris, D., & Muir, H. C. (1993). A psychophysiological approach to the assessment of work underload, *Ergonomics*, 36, 1035-1042.
- Comstock, J. R., & Arnegard, R. J. (1992). *The multi-attribute task battery for human operator workload and strategic behaviour research* (Tech. Memorandum No. 104174). Hampton, VA: NASA Langley Research Center.
- Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K. R. Boff, L. Kaufman, & J. P. Thomas (eds.), *Handbook of perception and human performance: Cognitive processes and performance*, Vol. II. (pp. 41/1 – 41/44). New York: Wiley.
- Hart, S. G., & Staveland, L.E. (1988). Development of the NASA-Task Load Index (NASA-TLX); Results of empirical and theoretical research. In P.A. Hancock and N. Meshkati (eds.), *Human mental workload* (139-183). Amsterdam: Elsevier.
- Hilburn, B., Jorna, P. G. A. M., & Parasuraman, R. (1995). The effect of advanced ATC automation on mental workload and monitoring performance: An empirical

- investigation in Dutch airspace. Proceedings of the 8th International Symposium on Aviation Psychology, Columbus, OH: The Ohio State University.
- Johannsen, G., Pfendler, C., & Stein, W. (1976). Human performance and workload in simulated landing-approaches with autopilot failure. In T. B. Sheridan & G. Johannsen (eds.), *Monitoring behaviour and supervisory control* (pp. 83-93). New York: Plenum.
- Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress (pp. 279-328). In D. L. Damos (ed.) *Multiple-task performance*. London: Taylor & Francis.
- Lee, J. D., & Moray, N. (1992). Trust control strategies, and allocation of function in human machine systems. *Ergonomics*, 35, 1243-1270.
- Liu, Y. R., Fuld, R., & Wickens, C. D. (1993). Monitoring behaviour in manual and automated scheduling systems. *International Journal of Man-Machine Studies*, 39, 1015-1029.
- Madigan Jr., E. F., & Tsang, P. S. (1990). A survey of pilot attitudes toward cockpit automation. Proceeding of the 5th Mid Central Ergonomics/Human Factors Conference, Norfolk: VA.
- Masalonis, A.J., Duley, J.A., & Parasuraman, R. (1999). Effects of manual and autopilot control on mental workload and vigilance during simulated general aviation flight, *Transportation Human Factors*, 1, 187-200.
- Moray, N. (1988). Mental workload since 1979. In D. Osborne (ed.), *International Review of Ergonomics*, 2, 123-150.
- Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors*, 33, 17-33.
- Parasuraman, R. (1987). Human-computer monitoring. *Human Factors*, 29, 695-706.
- Parasuraman, R. (1999). Automation and human performance. In W. Karwowski (ed.) *International encyclopedia of ergonomics and human factors*. New York: Taylor & Francis.
- Parasuraman, R., & Mouloua, M. (1996). *Automation and human performance: Theory and applications*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39, 230-253.
- Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced complacency. *International Journal of Aviation Psychology*, 3, 1-23.
- Parasuraman, R., Mouloua, M., Molloy, R., & Hilburn, B. (1996). Monitoring of automated systems. In R. Parasuraman & M. Mouloua (eds.), *Automation and human performance: Theory and applications* (pp. 91-115). Hillsdale, NJ: LEA.
- Singh, I. L., & Parasuraman, R. (2001). *Human performance in automated systems*. Indian Psychological Abstract and Reviews, 8:, 235-276.
- Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation-induced complacency: Development of the complacency potential rating scale. *International Journal of Aviation Psychology*, 3, 111-122.
- Singh, I. L., Molloy, R. & Parasuraman, R. (1997). Automation-induced monitoring inefficiency: Role of display location. *International Journal of Human Computer Studies*, 46, 17-46.
- Singh, I. L., Parasuraman, R., Deaton, J., & Molloy, R. (1993, November). Adaptive function allocation enhances pilot monitoring performance. Proceedings of the Annual Meeting of the Psychonomic Society. 34th Annual Conference, Washington, DC, USA.
- Wickens, C. D., & Kessel, C. (1981). Failure detection in dynamic system. In J. Rasmussen, & W. B. Rouse (eds.) *Human detection and diagnosis of system failures*. New York: Plenum.



- Wickens, C. D., & Kramer, A. F. (1985). Engineering psychology, *Annual Review of Psychology*, New York: Annual Review Inc.
- Wickens, C. D., Mavor, A.S. & McGee, J.P.(1997). *Flight to the future: Human factors in air traffic control*. Washington, DC: National Academy Press..
- Wiener, E. L. (1988). Cockpit automation. In E. L. Wiener & D. C. Nagel (eds.), *Human factors in aviation* (pp. 433-461). San Diego: Academic

**Indramani. L .Singh**, PhD, Reader, Department of Psychology, Banaras Hindu University, Varanasi.

**Hari Om Sharma**, Department of Psychology, Banaras Hindu University, Varanasi

**Anju L. Singh**, Department of Psychology, Banaras Hindu University, Varanasi

With Best Wishes from

## **Prasad Psycho Corporation**

**J-1/58, Dara Nagar,  
Varanasi - 221 001 (India)**

**B-39, Gurunanakpura,  
Laxmi Nagar,  
New Delhi - 110 092 (India)**

Phone: 011- 32903349, Fax: 011-41765277,  
Mobile: 09810782203