Applied Ergonomics 1989, **20**.3, 197–205

The effect of performance failure and task demand on the perception of mental workload

P.A. Hancock

Department of Safety Science and Human Factors Department, Institute of Safety and Systems Management,
University of Southern California, Los Angeles, USA

An experiment was conducted to investigate the effects of incipient performance failure on subjective workload response. Unknown to the participants, there were two levels of demand embedded in the PC-based flight simulation performance task. These were included to assess whether task-naive subjects could distinguish variations in load in the face of repetitive performance failure. Results confirmed that participants were able to distinguish this change in load level under such circumstances and that failure resulted in higher perceived workload than did successful performance. Within these differences were significant effects dependent upon subject gender. In general, female participants performed more poorly and rated workload higher than their male counterparts. It is concluded that the experience of failure presents a significant source of workload in systems operation.

Keywords: Human performance, mental workload, performance failure, task demand

Introduction

Mental workload response can be used in a number of ways to assist the designer and operator of contemporary systems (see Gopher and Donchin, 1986; Hancock and Meshkati, 1988; O'Donnell and Eggemeier, 1986). As a diagnostic tool, workload can help the ergonomist to distinguish between the efficiency of competing designs. Workload can also be used to provide insights into specific job characteristics (Kantowitz, 1987). Among the more promising areas of application is the on-line assessment of mental load imposed upon individual operators. This information is critical to the development and operation of certain adaptive systems based upon dynamic task allocation (Hancock and Chignell, 1987; 1989). Along with others (Damos, 1988; Derrick, 1988), we have sought to distinguish some of the characteristics that may help provide a prediction of load (Hancock, 1988a; Hancock and Chignell, 1988) and indicate the workload level which exceeds an individual operator's response capabilities. Such individual failure thresholds have yet to be distinguished unequivocally, even in controlled and easily manipulable laboratory tasks (but see Schlegel *et al,* 1988).

In the vast preponderance of existing studies on workload effects, a task is presented in which the subject engages in the performance of a difficult but achievable sequence of actions. As may be expected, workload is initially high but reduces with practice as the subject distinguishes effective response strategies (Hancock and Meshkati, 1988; Hancock *et al,* 1989). Subjects may perform more or less efficiently depending upon the measures used to assess performance, but generally error rates associated with such procedures are uniformly low. The present study employed a different approach in focusing upon incipient failure in a performance task as an added source of load. Unfortunately, in many real-world situations where extremes of workload may be experienced, no on-task practice may be available and the expectation of failure, like error (Hart and Bortolussi, 1984), may itself be a driver of the workload experience. As with many ergonomics endeavours, a modal environment for the evaluation of such phenomena is aerospace activities.

The present study sought to distinguish the effects of performance failure on task-naive participants, and to establish whether individuals could distinguish levels of task demand even during failure sequences. From previous work (Hancock, 1988a; Hancock *et al,* 1988), we have established that subject gender is a powerful influence on perceived workload in very low task demand conditions (see also Hancock, 1988b). To examine whether such effects are specific to low-load conditions, or if they generalise across the complete continuum of task demand, gender was also embedded in the current design.

Method

Task

In the present experiment it was necessary to choose a difficult but achievable task which was similar to some real-world performance demands. For this purpose, a commercially available PC-based F-16 flight simulation package (Falcon^{TM}) was selected. Using the keyboard control mode, subjects were required to perform a take-off sequence, a 270° left turn followed by an approach and landing at a cross runway. This

0003–6870/89/03 0197–09 $03.00

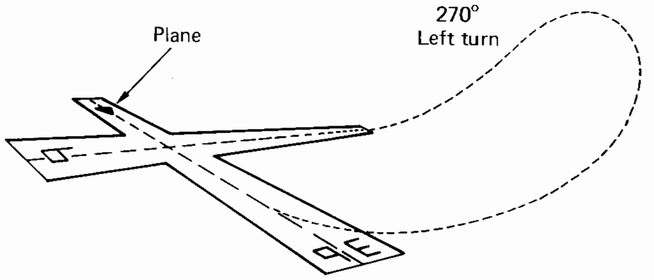

Fig. 1 Schematic illustration of the performance task. The subject was required to perform a take-off, a 270 degree left-turn followed by a landing on a cross runway

sequence is illustrated in Fig. 1. Subjects were requested to perform as quickly and accurately as possible but were instructed to place safety in landing as the primary performance goal. The subject was timed from the beginning of the start-up sequence to the termination of the flight, whether this was a crash or a successful return. Within the task itself were interpolated two differing levels of demand. In terms of the commercial software these were classed as First Lieutenant and Full Colonel levels, being the extremes of the range available.[1] We denoted them as Level 1 and Level 2, respectively.

Subjects

For the present experiment, there were 12 subjects, six male and six female, who were elicited as unpaid volunteers from the staff and student body of the University of Southern California. None of the subjects had any experience on the flight simulation task, and none of the subjects were frequent or skilled computer-based electronic games players. In so far as possible, subjects were age-matched across gender.

Workload assessment technique

The level of perceived workload was measured by the NASA Task Load Index (TLX). This assessment instrument has been described in detail by Hart and Staveland (1988). Briefly, the TLX presents a two-step sequence, one of which is completed prior to the experimental procedure and the second which typically immediately follows performance termination. In the first step, the participant compares the six identified dimensions of workload against each other in a pairwise manner to establish which dimension of each pair is perceived to contribute the greater source of load. Subjects have the definitions of each dimension in front of them during the comparisons and the process is self-paced. The dimensions are: mental demand (md), physical demand (pd), temporal demand (td), effort (ef), performance (op), and frustration (fr). A weighting is derived for each dimension depending upon the number of times it is selected against its companions in the 15 pairwise comparisons. The weights can vary between zero and five. In the former case a dimension has not been chosen as the greater source of load in comparison with any of its companions. In the latter case a dimension is chosen for each pair in which it appears.

The second step of the TLX process requires the subject to rate the unit of work they have just experienced. In the laboratory this is usually an event or trial. These events are rated on a 0–100 scale for each of the six dimensions. Each scale is broken into units of 10 by major subdivisions and units of five by smaller interspersed dividers. The descriptions of these dimensions and their rating sheet are shown in Figs. 2a and 2b, respectively. In the paper version, the subject marks a line through the level of load on each raw scale. On the computerised version the subject manipulates a cursor to mark the perceived load. These values represent the raw scores for each dimension. The raw scores are multiplied by their respective weights to derive weighted workload scores. The overall workload value is then given as the sum of all weighted scores divided by 15, the original number of pairwise comparisons.[2]

Design and procedure

In the present experiment a mixed design was employed. There was a single between-subject factor, which was gender, and two within-subject factors which were level of task demand and trial. Each subject performed the same number of trials and both levels of demand. Order of administration of task demand level was counterbalanced across the whole design and also within gender. Thus three subjects of each gender performed the easy-difficult sequence and three subjects of each gender performed the difficult-easy sequence. Subjects observed one instruction flight conducted by the experimenter to familiarise them with the keyboard controls and the exact flight procedure. At the termination of the demonstration flight, subjects were able to ask questions about the controls, as they were in between each flight trial. Two experimenters of differing gender administered the procedure to obviate the known potential for subject gender and experimenter gender interaction effects (Hancock, 1988a; Rumenik *et al*, 1977). There were 10 flight trials per subject. Responses were not given to questions concerning strategy for successful performance. The experimenter timed the duration of each flight in seconds from the start-up point until the termination in either a crash or successful landing. The former were coded as errors, the latter as successful trials, and these data were used for the landing status measure. At the end of each individual trial, subjects

---

[1] The software package utilised was Falcon^{TM}. The levels of demand are derived by routines embedded within the proprietary software. In the Index of the accompanying operator Manual these are described as follows:

"In the program, the calculations at Lieutenant level represent simple translations where throttle levels are directly equated with aircraft velocity. At the full Colonel level, designed to represent a more veridical F-16 simulation, additional inputs dependent upon simulated weight, angle of climb, drag and thrust-to-weight ratio are included."

As Falcon is freely available, it is possible to replicate the present work and to assert that the task demand increases with the load manipulations noted. Use of this software package does not imply endorsement by the author, the author's Institution or the Sponsoring agency.

[2] Both the pencil and paper and computerised versions (IBM compatible) of the Task Load Index (TLX) assessment of workload are available from the Human Factors and Rotorcraft Research Division, NASA Ames Research Center, Moffett Field, CA 94035, USA.

Rating Scale Definitions

Title	Endpoints	Descriptions
Mental demand	Low/High	How much mental and perceptual activity was required (e g, thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical demand	Low/High	How much physical activity was required (e g, pushing, pulling, turning, controlling, activating, etc)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal demand	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task element occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	Good/Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfield were you with your performance in accomplishing these goals?
Effort	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration level	Low/High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Fig. 2a Definition of the six NASA TLX dimensions

completed the computerised version of the TLX workload assessment questionnaire. Data from this procedure and the flight performance data were taken for analysis. This analysis was done using a Macintosh IIx microcomputer and Statview II software. The mixed analysis-of-variance procedures were performed on a Compaq 386/7-20 computer using Data Analysis System 2.02 software (Olofinboba *et al,* 1989).

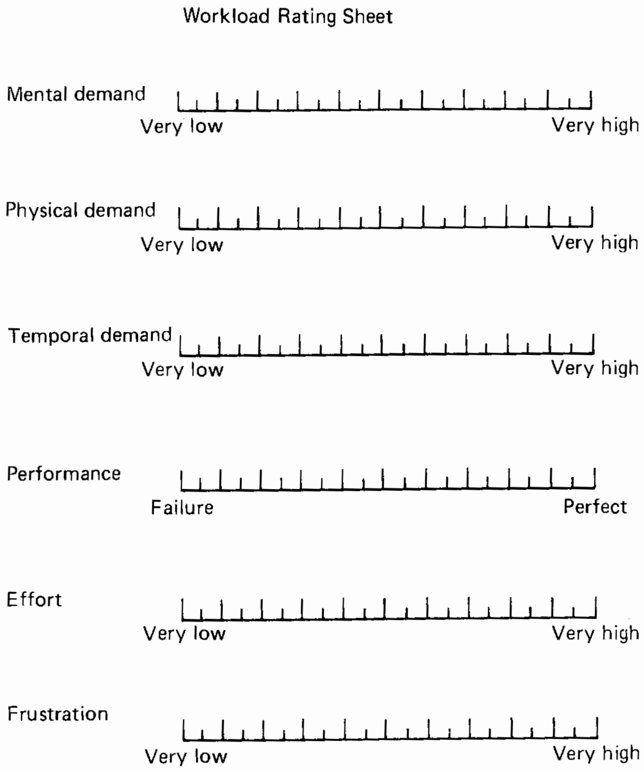

Fig. 2b Raw scoring scales for the six NASA TLX dimensions

Results

(1) Elapsed Flight Time

In the present experiment, there were two performance measures – flight time and landing status. While the original notion had been to consider a brief flight time as evidence of superior performance, it was clear that the opposite was true. Longer flight times, or the ability to keep the aircraft airborne, clearly denoted better performance in this task for these task-naive performers. The ANOVA table for flight time is given in Table 1. As can be seen, there were no interactive effects between the factors of operator gender, imposed task demand, and sequential trial. However, there were significant main effects for the within-subject factor of imposed task demand and the between-subject factor of gender. As illustrated in Fig. 3, increasing workload level resulted in a reduction of flight time of about one half. A similar difference can be seen in the main effect for gender where females exhibited a lower flight time than their male companions (see Fig. 4). There was no main effect for the trial factor. Such a finding implies no improvement or decrement in this demanding task during the first 10 attempts. It should be noted, however, that there were large inter- and intra-individual performance variations during these initial trials.

A concern for the present findings was the question of asymmetric transfer. This effect occurs when exposure to a previous condition influences performance in a subsequent condition but not vice versa. In the present case this can be distinguished as the influence of task demand order (i e, low

Table 1: Analysis of variance table for the dependent variable: Elapsed flight time

Source	Sum of squares	Mean squares	DF	F ratio	Prob if assump met	Prob with conserv adj
Between Subjects						
Gender	192593·25757	192593·25757	1	13·279	0·005	
Error	145040·64473	14504·06447	10			
Within Subjects						
TD	198481·80102	198481·80102	1	16·506	0·002	0·002
Gender x TD	30541·38040	30541·38040	1	2·540	0·142	0·142
Error	12047·78469	12024·77847	10			
Trial	1844·99808	466·74952	4	0·080	0·988	0·783
Gender x Trial	16564·2851	4141·07063	4	0·712	0·589	0·419
Error	232696·30219	5817·40755	40			
TD x Trial	32158·33361	8039·58340	4	1·288	0·291	0·283
Gender x TD x Trial	44391·52973	11097·88243	4	1·779	0·152	0·212
Error	249585·87700	6239·64693	40			

to high, versus high to low) on the 10 overall performance trials. Analysis of these factors on flight time revealed a significant interaction as illustrated in Fig. 5. This has been compressed into the mean values for the sequential trial blocks (i e, 1–5, and 6–10). As can be seen, the improvement from starting at the difficult level and subsequently transferring to the easy level is not equivalent to the performance change when going in the reverse order. The practical and methodological implications of such transfer are examined in detail in the discussion.

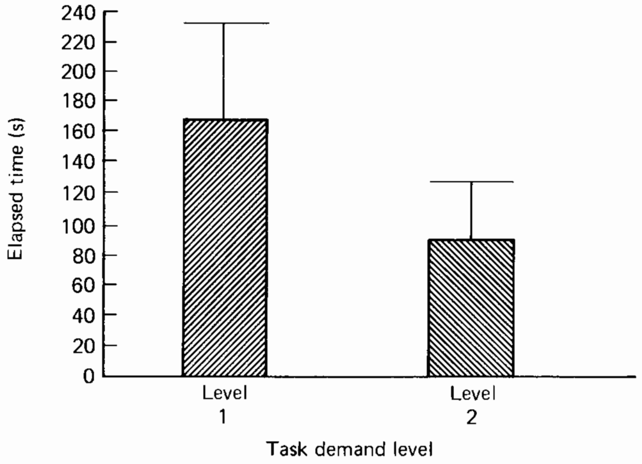

Fig. 3 Main effect for imposed task demand level versus elapsed flight time

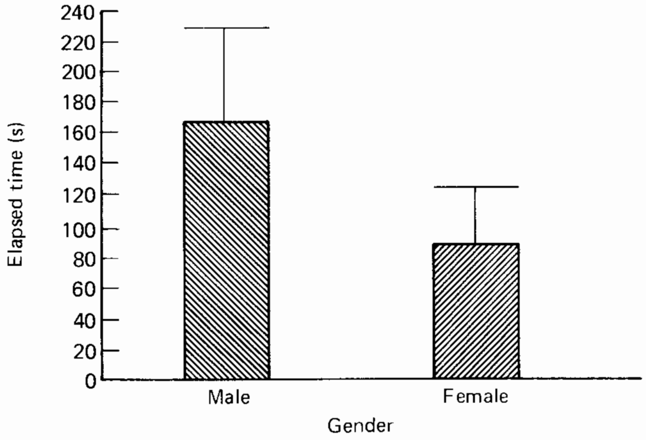

Fig. 4 Main effect for subject gender versus elapsed flight time

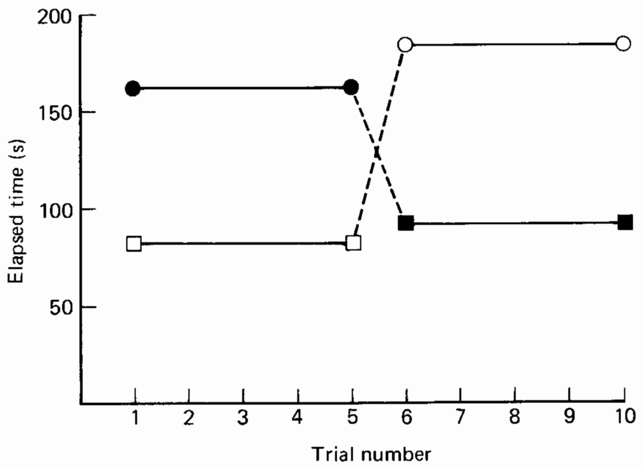

Fig. 5 Interaction between order of load presentation and trial for elapsed flight time. The differential effect on performance for load order low-to-high, compared with high-to-low, implies an asymmetric transfer effect. Open symbols respresent the first order from low-to-high load and closed symbols the reverse loading order. Circular symbols represent low load which is reflected in longer elapsed flight times, square symbols are high load with associated lower flight times

(2) Landing status

The effects for flight time were directly reflected in landing status. Using a Chi-Square analysis, there was a significant effect for imposed task demand (Chi-Square = 18·46, $p < 0·0001$). There were no successful landings at the higher workload level. There was also a significant effect for gender (Chi-Square = 14·13, $p < 0·0002$), where there was greater than a 10-fold difference in successful landings between male and female subjects. It should be reiterated that these were task-naive individuals and cannot be taken as evidence of any gender difference in task-competent performers. Simple observation reveals the interaction between operator gender and task demand level. As can be seen from Fig. 6, the single landing by one female at the lower demand level is matched against the 50% success rate for the males at the same level. It is important to emphasise again that this gender effect is for task-naive individuals. While no subject professed to being a frequent player of video games, it is possible that even passing experience may assist performance and that such experience is generally supposed to be more common among males compared to their age-matched female peers. However, the present data support a significant gender difference in these beginning performers.

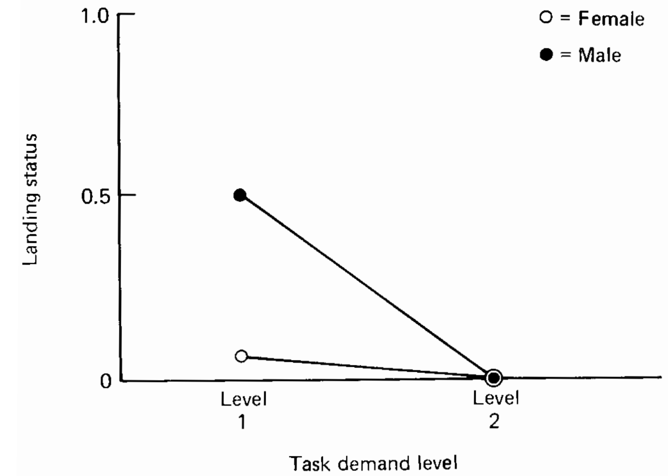

Fig. 6 Differential levels in percentage of successful landings at two imposed task loading dependent upon subject gender

Workload measures

The overall mean workload level and the raw workload ratings on each dimension were subjected to the same analysis as that for elapsed time, using the factors of task demand level, trial and performer gender. The ANOVA table for the overall workload measure is given in Table 2. Use of the ANOVA procedure has been applied previously to data derived from the NASA-TLX workload assessment technique (see Vidulich and Tsang, 1986). However, there is a concern for the scaling of equal intervals with this workload measure (for a discussion see Hart and Staveland, 1988), and because of this concern subsequent analysis used the probability with the conservative degrees of freedom adjustment for both raw and overall workload results. As can be seen from Table 2, for the overall workload there was a significant increase in perceived load which followed the task demand manipulation. As the manipulation was not communicated to the subjects, it is indicative that even under repeated failure conditions, subjects were able to register changes in externally imposed demand. This trend is illustrated in Fig. 7. The composition of these overall workload scores broken down by weighting and raw score on each of the six TLX dimensions are presented in Figs. 8 and 9. Fig. 8 gives the composition for the low task demand level (level 1), while Fig. 9 gives the comparable composition for the high task demand level (level 2).

There was also a significant interaction between subject gender and imposed task demand for the overall perceived workload value. This interaction, illustrated in Fig. 10, indicated that while female participants tended to rate each task demand level as comparable in terms of perceived load, the male participants discriminated between

Table 2: Analysis of variance table for the dependent variable: Overall workload

Source	**Sum of squares**	**Mean squares**	**DF**	**F ratio**	**Prob if assump met**	**Prob with conserv adj**
Between Subjects						
Gender	42·02017	42·02017	1	0·021	0:889	
Error	20444·09825	2044·40983	10			
Within Subjects						
TD	957·39252	957·39252	1	5·994	0·034	0·034
Gender x TD	894·73024	894·73024	1	5·601	0·039	0·039
Error	2096·37995	159·73349	10			
Trial	367·15732	91·78933	4	1·751	0·158	0·215
Gender x Trial	336·12559	84·03140	4	1·603	0·192	0·234
Error	2096·37995	52·40950	40			
TD x Trial	436·79492	109·19873	4	2·329	0·073	0·158
Gender x TD x Trial	442·93090	110·73272	4	2·362	0·069	0·155
Error	1875·57025	46·88926	40			

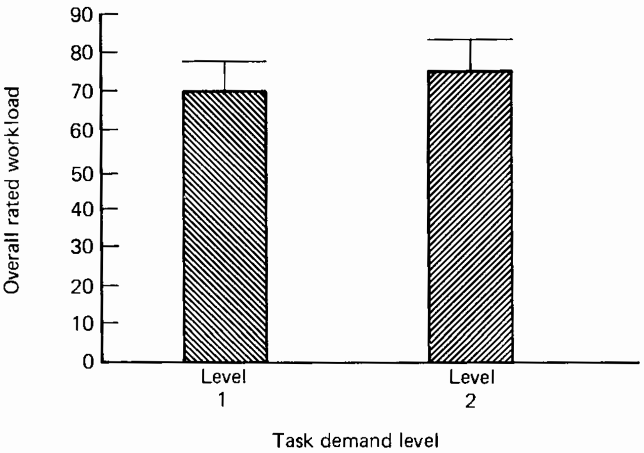

Fig. 7 Significant difference in overall perceived workload dependent upon the level of imposed task demand

differing levels. It is clear that such evaluations are closely related to the outcome on the performance measures, where males successfully accomplished the landing procedure 50% of the time in the low task demand condition, with no successful landings at the higher demand condition. Female participants experienced a similar failure rate at the high demand, but showed little success at the lower task demand level also. This differential effect also occurred in elapsed flight time where the relative change between the two workload levels is smaller for the female compared with the male participants. However, it should be noted that the latter interaction did not reach an acceptable standard for statistical significance (*F*[1,10] = 2·54, p = 0·14). An immediate question may then be raised as to whether the subjects were simply reacting to their perceived performance which itself is contingent upon externally imposed task demand.

Resolution of this proposition can be found in part in the performance dimension where subjects are specifically required to produce a score dependent on the perception of their own competence with respect to the task. As illustrated in Fig. 11, the raw rating on this dimension showed a significant main effect for workload level (*F*[1,10] = 6·93, p = 0·025). However, in contrast to what might have been expected, there was no interaction between gender and workload level, or any main effect for gender on this dimension. There was an indication of an interaction between imposed task demand and trial though. Effect level approached that for conventional significance (*F*[4,40] = 4·17, p = 0·069), with the correction for the conservative degrees of freedom adjustment. If analytic assumptions are met, this value is highly significant (p = 0·007). This tendency is illustrated in Fig. 12, and, as can be seen, the propensity to reduce perceived performance workload raw scores under the low task demand is contrasted with the consistently high performance workload raw estimates at the

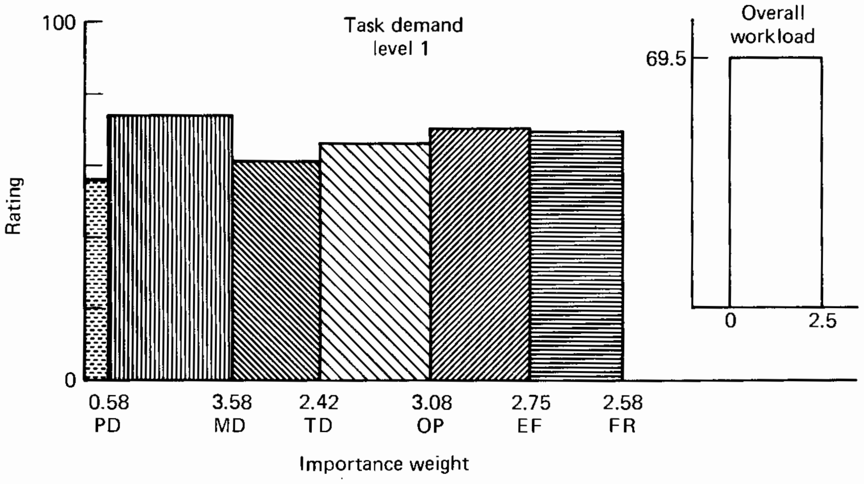

Fig. 8 NASA TLX overall workload rating and subdivision by rating dimension weight (horizontal axis) and raw rating (vertical axis) for task loading Level 1 (easy) in the flight simulation task

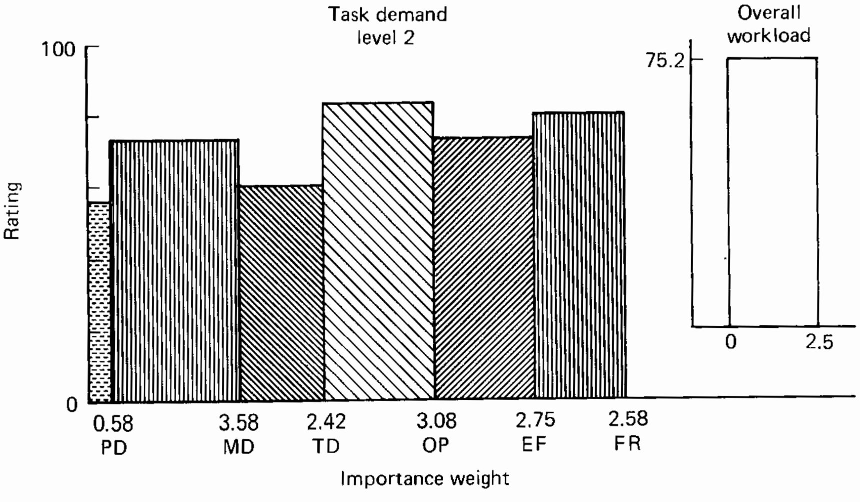

Fig. 9 NASA TLX overall workload rating and subdivision by rating weight (horizontal axis) and raw rating (vertical axis) for task loading Level 2 (difficult) in the flight simulation task

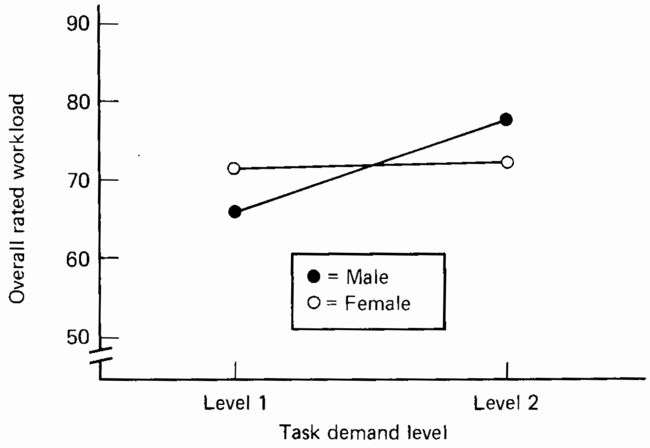

Fig. 10 Interaction between subject gender and task load level for the overall NASA TLX workload scale

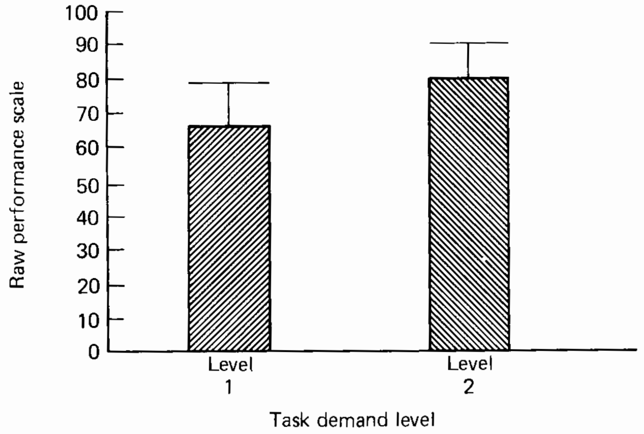

Fig. 11 Differential effect imposed task demand upon the TLX performance dimension

higher level of task demand. This trend is not contingent upon subject gender, as might be suspected from the results for overall workload, in that the three-way interaction is clearly not significant (p = 0·24).

The dimension which did present a pattern consistent with the performance results was perceived frustration. It should initially be recalled that failure was the predominant characteristic of this present task, and there were main effects for task loading and gender for each of the performance measures. In keeping with this pattern, female participants exhibited little change in reported frustration between task demand levels and expressed consistently high scores [demand level 1 (easy) = 80·5, demand level 2 (difficult) = 82]. However, the male subjects generated lower overall scores on this frustration dimension and showed substantial change between task demand levels [demand level 1 (easy) = 56·8, demand level 2 (difficult) = 78·2]. This interaction can be seen in Fig. 13. Females experienced a higher level of frustration than their male counterparts. However, these levels converged at the higher imposed demand level where both groups experienced a particularly high level of frustration. There is a potential ceiling effect involved in these trends as 100 is the maximum score on each raw scale. For the effort dimension there was an interaction that approached traditional levels of significance between task demand and gender (F[1,10] = 4·62, p = 0·057). While male subjects produced an increased effort score as demand level increased [demand level 1 (easy) = 65·3, demand level 2 (difficult) = 75·2], the female participants scored effort in the other direction [demand level 1 (easy) = 74·3, demand level 2 (difficult) = 70·5].

To confirm the difference in perceived workload between successful and unsuccessful performance, an analysis was conducted of the respective scores. As illustrated in Fig. 14, there was a highly significant difference between overall workload scores, with the higher workload perceived in the unsuccessful performance condition. This analysis is reasonable given the time order of the production of response scores – i e, workload evaluation followed the termination of performance. Ideally, future measures of workload would occur during performance. However, the sequence of causality must be examined carefully. Does the higher workload induce performance failure and is then subsequently reported as an accurate reflection of momentary load, or does failure cause the perception of high workload response in and of itself? While the latter

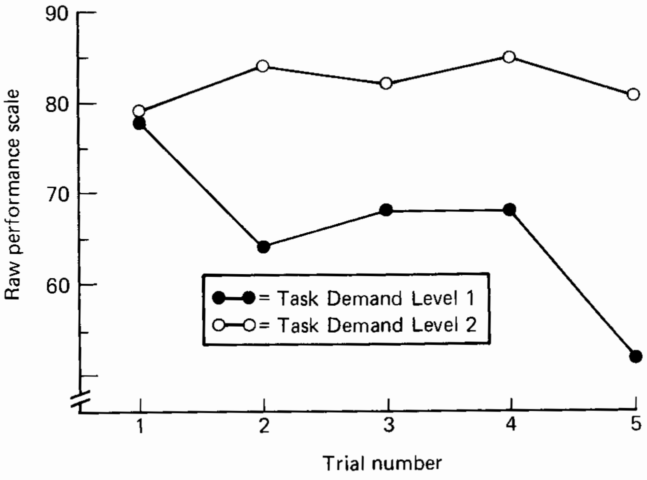

Fig. 12 Indication of an interaction between imposed task demand and trial in the TLX performance dimension

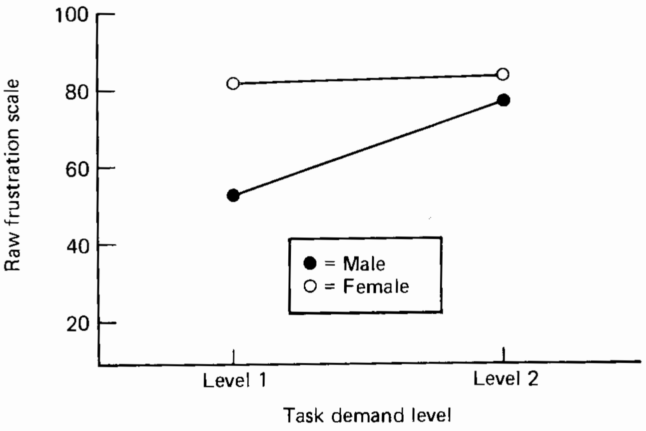

Fig. 13 Interaction between gender and task demand as represented in the TLX raw frustration dimension

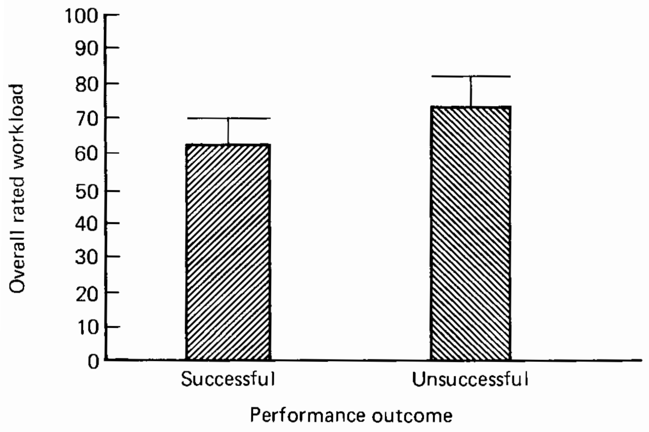

Fig. 14 Difference in perceived workload level dependent upon performance outcome

interpretation is adopted here, there is a clear need to investigate this potential feedback effect between momentary workload and progressive performance failure in actual working conditions.

Discussion and summary

We are used to studying human performance that is largely successful and where associated error rate is small. Similarly, in studying human-machine systems, predominant interest is given to activity during dynamically stable states of operation. However, with the increasing number of systems with zero error tolerance, we need to know more concerning the progress toward and recovery from differing states of failure. This necessity applies to an understanding of both human capabilities alone and when the operator works in conjunction with a complex system. The present results indicate that the experience of failure is a driver of workload. Preferably, we would like to measure workload during the task itself, rather than elicit response following termination with the associated potential problems that this form of assessment may bring (see Eggemeier, 1988; Eggemeier *et al,* 1984). However, the general point is that performance and workload during the dynamic instability that represents progressive failure is not well understood. Indeed, we need to know through what paths of action recovery may occur, and what role operator workload plays in the recovery of stable operational states.

One finding in the present work was asymmetric transfer between task load conditions. Essentially, the prior experience of one level of load did not transfer equally to performance at the other load level. The experimental psychologist may look upon such transfer as an artifact of the particular design chosen (Damos, 1985; Poulton, 1982). Indeed, it can be argued that such transfer is an inherent property of within-subject designs in general. This question is one of methodological pertinence and should rightly be dealt with elsewhere. However, it is clear that the level of impact of such transfer effects are contingent upon the type of performance task under consideration. In the present data set, we see that transfer from the low to high task demand condition (the filled symbols in Fig. 5) resulted in a reduction in flight time of approximately 60 s from 153 s to 92·4 s. This difference has to be compared to the 100 s difference from 184·9 s to 82·8 s for the reverse order from high to low demand. Using the first five trials in each order as the baseline it is clear that prior practice on the low load level improved performance on the high load by approximately 10 s (i e, 92·4 s to 82·8 s). For the reverse order, prior to practice at the high load level resulted in an approximate 30 s increase in flight time (i e, 184·9 s to 153 s). Unfortunately, in this task – as with many other performance tasks – we have no way to quantify the difference in the levels of difficulty other than the performance outcome itself. In the present case, if we assume a linear model of demand then the difficulty of the high demand level is three times that of the low demand level (i e, 30 s difference compared with 10 s difference). This would assume no actual asymmetric transfer, and a direct relationship between performance time and task difficulty. Neither assumption can be adequately supported at the present time. For the practising ergonomist there are additional and somewhat different concerns from the purely methodological questions. For instance, depending upon the level of demand expected during normal performance, we may order practice in one manner whereas catering for expected load during incipient failure demands the imposition of practice loads in a different order. So asymmetric transfer may prove a practical consideration in training procedures, which is of greater interest to the applied scientist.

The above results indicate a number of effects with respect to mental workload and its generation. First, despite the fact that subjects repeatedly encountered performance failure in the present study, they were able to distinguish a hidden task load manipulation embedded into the design. While this observation holds across the participants tested, division on the basis of gender belies the general interpretation. It may be argued that while the male subjects were able to distinguish variation in imposed task load, the uniformly high scores by the female participants did not reveal such an ability. However, as noted earlier, a number of subjects reported the maximum scores on a number of dimensions, and as this was particularly the case for female subjects, then their apparent insensitivity may reflect a ceiling effect in the workload measure. Second, in comparison with several differing investigations that have used the TLX workload scale, the present overall scores represent some of the highest that have been recorded. This is in comparison with prolonged second order tracking (Hancock *et al,* 1989), difficult display discrimination (Gluckman *et al,* 1988) and numerous other performance tasks (Hart and Staveland, 1988). Third, there were significant effects for subject gender which were consistent across the performance and workload measures. The simple explanation that subjects rated workload according to their own perceived performance was not affirmed by examination of the raw scores on the performance dimension which addressed this postulate. If anything, the frustration dimension provided a more sensitive measure of this loading factor, as the interactions associated with load level and gender in task performance were replicated within this dimension.

Contemporary models of human performance are largely founded on the relatively sterile findings of the experimental laboratory. In general, such models do not capture the qualitative difference in operator reaction that occurs under highly stressful conditions (although see Hancock and Chignell, 1985; Hancock and Warm, 1989; Hockey, 1986; and Sanders, 1983). Further, they often neglect the whole spectrum of energetic characteristics of response which are so influential in real-world operations (Hockey *et al*, 1986). The present findings suggest that the experience of failure influences perceived workload on a difficult but achievable performance task. It is in precisely these conditions that information concerning operator and system response becomes most valuable. The practising ergonomist needs to recognise the symptoms which indicate deviation from normal operational states of both the system and its operator. Also, it is important to establish the sequence of actions which enable recovery to stability. Provision of such knowledge to the practitioner depends upon considerable effort to understand workload response during the process of failure as it occurs in the real-world environment.

Acknowledgements

This research was supported by Grant NCC 2-379 from the National Aeronautics and Space Administration (NASA), Ames Research Center. Sandra Hart was the technical monitor for the grant. The position presented here does not necessarily represent that of the named agency. The author would like to thank Dr John Wilson and two unknown reviewers with their helpful comments on an earlier version of this paper. The comments of my colleagues, Max Vercruyssen, Mark Chignell and Diane Damos, have been most helpful in revising this work. I would like to thank Andrew Chu for his production of the illustrations for this work.

References

Damos, D.L. 1985, *Human Factors,* 27, 409–421. The effect of asymmetric transfer and speech technology on dual-task performance.

Damos, D.L. 1988, Individual differences in subjective estimates of workload. In: P.A. Hancock and N. Meshkati (Eds). Human mental workload, North-Holland, Amsterdam. 231–237.

Derrick, W. 1988, *Human Factors,* **30**, 95–110. Dimensions of operator workload.

Eggemeier, F.T. 1988, Properties of workload assessment techniques. In: P.A. Hancock and N. Meshkati (Eds). Human mental workload. North-Holland, Amsterdam. 41–62.

Eggemeier, F.T., Melville, B.E., and Crabtree, M.S. 1984, *Proc Human Factors Soc,* **28**, 954–958. The effect of intervening task performance on subjective workload ratings.

Gluckman, J.P., Warm, J.S., Dember, W.N., Thielmann, J.A., and Hancock, P.A. 1988, Subjective workload response to simultaneous and successive vigilance tasks. Paper presented at the Annual Meeting of the Psychonomic Society, Chicago, IL.

Gropher, D., and Donchin, E. 1986, Workload: An examination of the concept. In: K. Boff, L. Kaufman, and J.P. Thomas (Eds). Handbook of perception and human performance. Wiley, New York, 41: 1–49.

Hancock, P.A. 1988a, The effect of gender and time of day upon the subjective estimate of mental workload during the performance of a simple task. In: P.A. Hancock and N. Meshkati (Eds). Human mental workload. North-Holland, Amsterdam. 239–250.

Hancock, P.A. 1988b, Tasks as stresses: The case of vigilance. In: Proc Psychophysics Symp. Stirling, Scotland.

Hancock, P.A., and Chignell, M.H., 1985, The principle of maximum adaptability in setting stress tolerance standards. In: R. Eberts and C.E. Eberts (Eds). Trends in Ergonomics/Human Factors II. North-Holland, Amsterdam. 117–125.

Hancock, P.A., and Chignell, M.H. 1987, Adaptive control in human-machine systems. In: P.A. Hancock (Ed). Human factors psychology. North-Holland, Amsterdam. 305–345.

Hancock, P.A., and Chignell, M.H. 1988, *IEEE Trans on Systems, Man and Cybernetics,* **18**, 647–658. Mental workload dynamics in adaptive interface design.

Hancock, P.A., and Chignell, M.H. (Eds). 1989, Intelligent interfaces: Theory, research and design. North-Holland, Amsterdam.

Hancock, P.A., and Meshkati, N. (Eds) 1988, Human mental workload. North-Holland, Amsterdam.

Hancock, P.A., Robinson, M.A., Grose, E., Chu, A., Hansen, D., and Vercruyssen, M. 1989, *Proc Human Factors Soc,* in press. Effects of training on tracking and subjective workload.

Hancock, P.A., Rodenberg, G.J., Mathews, W., and Vercruyssen, M. 1988, *Proc Human Factors Soc,* 32, 857–861. Estimation of duration and mental workload at differing times of the day by males and females.

Hancock, P.A., and Warm, J.S. 1989, A dynamic model of stress and sustained attention. Report TRL-NASA-8803, University of Southern California, Los Angeles, CA.

Hart, S.G., and Bortolussi, M.R. 1984, *Human Factors,* 26, 545–556. Pilot errors as a source of workload.

Hart, S.G., and Staveland, L.E. 1988, Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: P.A. Hancock and N. Meshkati (Eds). Human mental workload. North-Holland, Amsterdam. 139–183.

Hockey, G.R.J. 1986, Changes in operator efficiency as a function of environmental stress, fatigue, and circadian rhythms. In: K.R. Boff, L. Kaufman, and J.P. Thomas (Eds). Handbook of perception and human performance. Wiley, New York, 44: 1–49.

Hockey, G.R.J., Gaillard, A.W.K., and Coles, M.G.H. (Eds). 1986, Energetics and human information processing. Nijhoff, Dordrecht.

Kantowitz, B.H. 1987, Mental workload. In: P.A. Hancock (Ed). Human factors psychology. North-Holland, Amsterdam, 81–121.

O'Donnell, R., and Eggemeier, F.T. 1986, Workload assessment methodology. In: K.R. Boff, L. Kaufman, and J.P. Thomas (Eds). Handbook of perception and human performance. Wiley, New York, 42: 1–49.

Olofinboba, A., Vercruyssen, M., and Wheeler, W.F. 1989, DAS 202: Data analysis system for IBM compatible microcomputers. Computer Software, Psy-Med Associates, 2329 Sonoma, Honolulu, HI 96822.

Poulton, E.C. 1982, *Psychol Bull,* **92**, 673–690. Influential companions: Effects of one strategy on another in the within-subjects designs of cognitive psychology.

Rumenik, D.K., Capasso, D.R., and Hendrick, C. 1977, *Psychol Bull,* **84**, 852–877. Experimenter sex effects in behavioural research.

Sanders, A.F. 1983, *Acta Psychologica,* **53**, 61–97. Toward a model of stress and human performance.

Schlegel, B., Schlegel, R.E., and Gilliland, K. 1988, *Proc Int Ergonomics Assoc,* **10**, 475–477. Determining excessive mental workload with the subjective workload assessment technique.

Vidulich, M.A., and Tsang, P.S. 1986, *Ergonomics,* 29, 1385–1398. Techniques of subjective workload assessment: A comparison of SWAT and the NASA-Bipolar methods.