TR 90019

TR 90019

(2)

ROYAL AEROSPACE ESTABLISHMENT

Technical Report TR 90019

March 1990

A Subjective Rating Scale for Assessing Pilot Workload in Flight: A Decade of Practical Use

by

A. H. Roscoe
G. A. Ellis

Procurement Executive, Ministry of Defence
Farnborough, Hampshire

UNLIMITED

UNLIMITED

ROYAL AEROSPACE ESTABLISHMENT

Technical Report 90019

Received for printing 5 March 1990

A SUBJECTIVE RATING SCALE FOR ASSESSING PILOT WORKLOAD IN FLIGHT: A DECADE OF PRACTICAL USE

by

A. H. Roscoe*

G. A. Ellis**

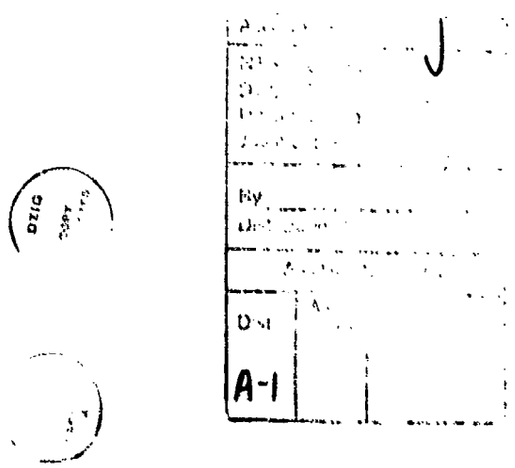
SUMMARY

Despite the many techniques developed for evaluating pilot workload in flight, subjective assessment by experienced pilots is still the most reliable method by far. This report describes the design and development - with the help of practising test pilots - of a ten-point rating scale. The scale uses a decision tree similar to that used by the Cooper-Harper Handling Qualities scale, and is based on the concept of spare capacity. Examples are given of its use by a large number of pilots in various flight trials and workload studies.

Departmental Reference: FM 6

* Formerly of RAE, Bedford, and now with Britannia Airways
** Formerly of RAE, Bedford, and now with British Aerospace

LIST OF CONTENTS



1 INTRODUCTION

In 1969 the increasing importance to flight safety of changes in levels of pilot workload generated by new operating techniques such as vertical take-off and landing (VTOL), low visibility landings, and reduced noise approaches, then being evaluated at the Royal Aircraft Establishment at Bedford, resulted in a greater interest in assessing workload during flight testing. Unstructured pilot opinion recorded during flight, or, more often, after flight was the accepted method of evaluating workload at this time. The possibility of obtaining misleading information because of bias or of pre-conceived notions about workload levels - a recognised problem associated with subjective techniques - resulted in a programme aimed at developing a complementary but more independent measure. Following a detailed survey of available techniques, monitoring of pilot's heart rates was selected for further evaluation.

After some 5 years, during which considerable experience was gained of using test pilot's heart rates to support their opinions of workload, it was decided to improve the method of obtaining subjective assessments by employing a specially designed rating scale. A search for a suitable scale for use in flight was unsuccessful; most research on subjective assessment by pilots had been concerned with aircraft handling qualities[1-3]. Although some scales such as that designed by Cooper and Harper[3] have sometimes been used to rate workload they are not ideal for this purpose. As Gerathewhol[4] pointed out: "... subjective pilot ratings of handling qualities, as accurate as they may be in regard to control desirability or difficulty, do not contribute to workload determinations, since they are only loosely connected to task demands and pilot response". The decision was made, therefore, to design and develop a workload rating scale at RAE Bedford 'on the back' of current flight testing and with the help of practising test pilots.

As well as designing a pilot workload rating scale it seemed sensible to define workload. A review of the literature revealed a plethora of definitions based mostly on workload as a set of flight task demands, as the effort required to satisfy those demands, or as the results of that effort - performance. Many of the definitions appeared complicated and/or unrealistic in the context of real flight.

Using a questionnaire, Ellis and Roscoe[5] obtained the views of some 330 military and airline pilots and concluded that more than 80% of professional pilots think of workload in terms of effort. This is also an interpretation that agrees well with the influence of such individual factors as natural ability,

training, and experience on the piloting task. There is evidence that the failure of pilots to perceive the demands of the flight task correctly has been a causative factor in several accidents; it also seems likely that workload levels tend to be determined by how a pilot assesses the flight task. With these findings in mind Ellis and Roscoe proposed that a slight modification to the definition of workload used by Cooper and Harper in the introduction to their Handling Qualities Rating Scale[3] would be appropriate, namely: "Pilot workload is the integrated mental and physical effort required to satisfy the perceived demands of a specified flight task".

2 THE 'BEDFORD' WORKLOAD SCALE

2.1 Design and development of the 'Bedford' scale

The initial objective was to design an interval scale that could be used to give ratings in flight during - or immediately following - highly demanding piloting tasks, and that would result in absolute values of workload.

As Ellis[6] observed: "The use of rating scales results in the allocation of a numerical value to the quantity that is being measured. Not unnaturally, researchers wish to use statistical and mathematical processes on the numbers so obtained, and so most of the rating scales that have been devised have been intended to be linear". However, McDonnell[7], in discussing the rating of aircraft handling qualities, referred to the difficulty of achieving linearity with ordinal and adjectival scales. And Hess[8] later commented: "The majority of rating scales in existence have two things in common: they are both ordinal and adjectival in nature". Furthermore, the results of various other laboratory studies[9,10] aimed at developing non-adjectival linear rating scales were not encouraging.

It became obvious that an ordinal, adjectival rating scale would be most appropriate for development at a flight test centre lacking facilities for laboratory experiments. An ordinal scale for aircraft handling qualities, the Cooper-Harper scale[3], was already developed and established; it was easy to use and was widely accepted amongst test pilots and engineers. It therefore made good sense to try to design the workload scale using a similar design of decision tree with appropriate descriptors.

The first design, a nine-point scale, used descriptors based on 'effort' such as: Pilot effort not a factor for desired performance - rating 1; desired performance requires moderate pilot effort - rating 3; adequate performance requires extensive pilot effort - rating 5; intensive pilot effort is required to

retain control - rating 8; and finally, control will be lost during some portions of required operation - rating 9.

At first it was not obvious whether the scale should be absolute or comparative. In other words, should it try to cover all possible workload levels in all flying tasks? Or should it have the more limited aim of acting as a comparative measure between the workload experienced and that which could be considered normal or reasonable for the task in hand? It was therefore decided to construct both types of scale, and then to decide which of them would be the more appropriate in various circumstances.

An interesting finding from the questionnaire study by Ellis and Roscoe[5] was that pilots find it convenient to think in terms of 'spare capacity' when considering their levels of workload. What other relevant but secondary tasks can be taken on in addition to the primary flight task? For example, when the primary task is an instrument approach in bad weather how much spare capacity does one consider is available for monitoring the actions of the other pilot, looking outside the cockpit, listening to the radio etc? The higher the workload generated by the primary task the less capacity there is for these secondary tasks. Pilots seem to find it a relatively simple matter to judge how much more they could do even if there is no requirement to do so.

These findings suggested that descriptors incorporating the concept of spare capacity would be of greater value than reference to effort. The scales were also extended from nine to ten ratings.

In addition to the concept of spare capacity the 'absolute' scale also referred to arousal, time, and fatigue. Those new descriptors included:

Low workload, plenty of time and capacity to complete all tasks at a moderate arousal state, level of effort could be maintained for several hours: - rating 2.

Moderate workload, all primary and secondary tasks within pilot capacity, but fairly high arousal state needed; tiring and fatigue likely after 1-2 hours: - rating 3. Very high workload, only more important secondary tasks completed and then only infrequently: - rating 6.

In view of the present concern about underarousal and underload it is worth noting that the lowest workload descriptor in this scale read: Very low workload, few tasks for the time available, some risk of boredom: - rating 1. This was shortly amended to: Workload too low, too much spare capacity, danger of complacency.

During construction of the scales it quickly became clear that any 'absolute' scale that attempted to include the whole spectrum of workload levels experienced by pilots would be too coarse to be practical. Certain tasks such as gun aiming or landing in adverse weather were always concentrated within a few seconds and were always high workload. Others, such as en-route flying, could be sustained for many hours. To place them on the same scale was of very limited value. Also, pilots comments and ratings of workload made during various flight tests in different types of aircraft, showed the difficulty of obtaining absolute values. It was found that pilots liked to compare their workload to some form of baseline, usually to previous experience.

Effort was therefore concentrated upon developing a comparative scale that would help to answer the important and practical question of whether the workload is appropriate for the primary task under consideration. Subsequent experience of using the scale has proved this decision to have been correct (see later).

The need for concise descriptors in an adjectival rating scale was highlighted during the evaluation of the absolute scale in flight. It soon became apparent that the introduction of the additional factors of arousal and fatigue complicated the scale unduly. After some development in flight and further discussions with pilots the present descriptors were introduced (Fig 1); these were readily accepted and in a short time considered by Bedford test pilots to be quite adequate for the purpose of rating workload.

2.2 Description of the scale (Fig 1)

The pilot starts his decision-making process at the bottom left corner of the decision tree, which consists of three questions requiring yes or no answers, in order to proceed to the descriptions of different levels of workload. The descriptors are of increasing levels of workload associated with ratings of 1 to 10. Half ratings are allowed thereby increasing the sensitivity of the scale, this became particularly desirable at the lower workload levels. Originally half ratings between the 'decision' groups were not sought but as many pilots seemed to find it difficult to decide between 'yes' and 'no' for "Was workload satisfactory without reduction?" a rating of 3½ became acceptable.

It is most important that the flight task to be rated should be well defined and/or the period of time over which the assessment is to be made stated with reasonable precision. The workload being assessed is that involved in the execution of the primary task; any additional tasks - such as monitoring other crew members - must be included as part of the pilot's spare capacity.

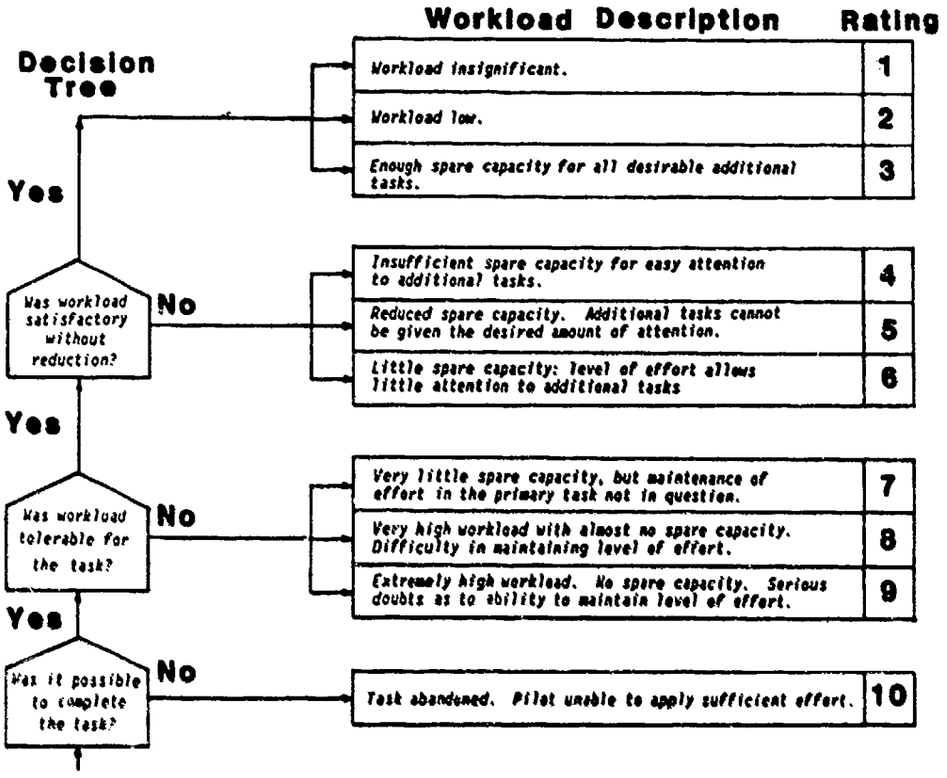


Fig 1 Pilot workload rating scale (for a specified piloting task)

3 PRACTICAL EXPERIENCE WITH THE 'BEDFORD' SCALE

The final version of the rating scale was first used to assess pilot workload during the HS Harrier ski-jump take-off trials[11]. This trial assessed the advantages of using an inclined ramp to improve the take-off performance of ship-borne Harrier VTOL combat aircraft. The aircraft is accelerated on to the ramp from a short run - typically 50-100 metres - with nozzles rotated rearwards. At the top of the ramp, and on the point of becoming airborne, the nozzles are rotated downwards to a pre-set angle. As conventional flying speed is approached the nozzles are gradually rotated to the aft position again.

The Cooper-Harper scale was used to assess handling qualities and the Bedford scale to estimate workload levels, the latter being augmented by recording the pilot's heart rate.

Eleven pilots rated their workload and had their heart rates recorded during ramp take-offs; the ramp angle was increased in steps from 6° to 15° over the period of the trial. Workload ratings and heart rates showed good agreement; and both ratings and heart rates confirmed that workload levels were not increased for greater ramp angles nor for night take-offs. These workload indicators also demonstrated that levels of workload are higher during the more conventional short take-offs (for this aircraft) from a runway.

The workload rating scale was used extensively during a trial to evaluate Economic Category 3 approach and landings[12]. Pilots' heart rates were again recorded to complement their ratings of workload. The technique involved an autopilot coupled approach to a decision height of 50 ft for the HS 748 and 60 ft for the BAC 1-11 aircraft at which height the autopilot was disconnected for a manual landing if the runway lights were seen. If the lights were not seen by decision height a go-around was made.

In addition to rating the final approach with autopilot, and the manual landing, ratings were given for the very short term workload associated with making the decision.

In late 1982 the Bedford scale and pilots' heart rate responses were used to assess workload in flight during crew complement certification of the BAe 146[13]. Post flight questionnaires complemented the in-flight data obtained from the three teams (of two pilots) who each flew three days of intensive flight schedules around a circuit of three high intensity airports (London - Heathrow, Paris - Charles de Gaulle, and Amsterdam - Schiphol).

Workload ratings were obtained from both pilots and from an experienced flight observer, on verbal request and light signal from the exercise controller, by means of small keyboards fitted to the control column and to the observer's clip board. Ratings, which were plotted automatically onto the heart rate plots at the time of the request, were requested according to a predetermined plan: requests were more frequent during high workload phases of flight such as the take-off and initial climb, the approach and landing, and when simulated in-flight failures and emergencies occurred.

Pilots were instructed in the use of the rating scale before the exercise started and, in particular, were asked to consider their workload for the previous 30 s. All six pilots and most of the flight observers found the scale easy to use; there was no evidence that giving ratings in flight intruded into the piloting task and only rarely was a rating delayed by the flight demands. Half ratings were not used during this trial and though the sensitivity of the

scale was consequently reduced it did not appear to influence its value for certification purposes. In fact, the ratings obtained during the trial were considered, overall, to be of considerable value; and there was also surprisingly good agreement between ratings given by pilots and those given by most of the flight observers.

There was a reasonably good relationship between pilot's ratings and their heart rate responses; disagreements seemed to be due mostly to the failure of the pilot to rate the entire period under review.

Lidderdale[14] used the Bedford scale and heart rate recordings to assess crew workload (pilot and navigator) during the evaluation of low-level high-speed flight in a supersonic tactical fighter aircraft. He reported that "... the aircrew understood the scale readily and whilst it was sufficiently comprehensive to cover all circumstances it was easy to remember and small enough to be carried on the flying suit knee-pad". At pre-planned times the navigator, having recorded his own rating of workload, would request a rating from the pilot. Both pilots and navigators reported little difficulty in giving ratings in flight. A post-flight assessment technique, based on the Analytical Hierarchy Process[15], was used to analyse paired comparisons. Lidderdale, in reporting a high correlation between in-flight ratings and the post-flight assessments, observed that the former technique was easier to use and more practical for use in an operational trial (personal communication).

The Bedford scale was used with success by Muir and Elwell[16] to assess workload in army helicopter pilots engaged in various flight tasks; and by Barnes[17] during an investigation into the levels of workload and operating conditions experienced by helicopter pilots involved in North Sea oil platform flights.

Roscoe and Grieve[18] used the Bedford scale together with heart rate responses to compare the levels of pilot workload generated by the advanced technology Boeing 767 with those generated by the earlier Boeing 737. This Britannia Airways study was carried out during routine passenger flights in Europe during which line-pilots found no difficulty in giving ratings at the end of the particular flight phase or sub-phase of interest. An experienced flight observer, using the Bedford scale, also rated the different flight tasks. Five pilots participated in the study, three were monitored on the B737 and then, after converting to type, on the B767, two pilots were monitored on both aircraft as they alternated every six months between B737 and B767. Both workload ratings and heart rate responses showed that the handling pilot's levels of workload are

lower on the 767 than on the 737. These measures have also distinguished quite clearly between the different levels of workload associated with flying the 767 in different modes: hand-flying with raw information, hand-flying with flight director integrated with the flight management system (FMS), and with autopilot and autothrottle.

An extension to this study is presently under way in which workload levels generated by flight failures, emergencies, and abnormal operating conditions in the 767 are being assessed in a flight simulator. Both ratings and heart rates for normal flight in the aircraft and simulator were used successfully to demonstrate the value of the simulator for this type of investigation[19].

Practical experience of using the workload rating scale in several flight trials showed it to be markedly better than previous methods of obtaining pilot opinion, and though it lacked some sensitivity - especially at the lower levels - it was becoming well accepted by pilots and by research scientists alike. Nevertheless, in 1983 it was decided to carry out a series of flights to demonstrate the ability of the rating scale and of heart rate responses to distinguish between four short flight tasks having, theoretically, three different levels of difficulty. A HS125 twin 'business-jet' was used for the trial in which 12 experienced pilots flew a total of 15 sequences. Each sequence consisted of :

(a) A 360° turn in 2 min at constant altitude, IAS, and rate of turn.

(b) A 360° turn in 2 min with a simultaneous loss of 2000 ft in altitude at a constant IAS and rate of turn.

(c) A 360° turn in 2 min with a simultaneous 2000 ft altitude loss followed by a reverse 360° turn in 2 min with a simultaneous gain of 2000 ft at a constant IAS and rate of turn.

(d) A 360° turn in 2 min with a simultaneous altitude loss of 2000 ft and speed reduction of 100 kn.

Each sequence was flown at a safe height in clear airspace. At first performance was monitored by the non-handling pilot but as it soon became obvious that each pilot was determined to perform well in the presence of a colleague this was discontinued. It was the intention to vary the order in which the tasks were flown - but, operationally, it was much more convenient to fly each sequence in the order 1-4 with the first task being repeated at the end of the sequence. Heart rates were recorded throughout each sequence and workload ratings were requested after each flight task. Three pilots flew the sequences twice. Those pilots not current on the 125 were given at least 30 min familiarisation before

being asked to rate the tasks; similarly, those pilots unfamiliar with the rating scale were given a full briefing beforehand.

Results were highly encouraging and demonstrated that for 10 of the 12 pilots both workload ratings and heart rate responses were able reliably to distinguish between the different tasks. In addition there was a very good agreement between ratings and heart rates for six of the 10 pilots and reasonably good agreement for the other four.

The mean ratings and range for each flight task were as follows:

Task	Mean Rating	Range
1	4.8	3-6
2	6.1	4-7
3	7.1	5-8
4	7.0	4-8
1	5.0	3-6

On theoretical grounds it was questionable whether task - 3, which lasted twice as long as the other three, would be rated lower than task - 4. In the event, task - 3 was rated higher than task - 4 on three occasions, lower on five, and the same on seven occasions. Although pilots were asked to rate the entire task, half of the pilots gave two ratings for task - 3.

Overall, there was no evidence of learning and reduced workload between task - 1 flown at the start and at the end of the sequence. Similarly, three pilots who flew the sequence on two separate sorties did not show any evidence of workload reduction with increasing familiarity.

4 DISCUSSION

The Bedford scale has now been in use for more than 10 years and though designed primarily for use by test pilots has been used successfully by military pilots, civil helicopter pilots, and airline pilots. The idea of 'spare capacity' appeals to most pilots who report that it helps them to arrive at an appropriate rating with relative ease, this is particularly valuable when first using the scale. In practice, pilots appear to become familiar with the scale remarkably quickly - most pilots then seem to think only in terms of numbers without reference to the actual decision tree.

The advantage of being able to use a rating scale during, or shortly after, a demanding flight task has been demonstrated many times during the 10 years - especially during long flight sectors requiring many ratings. Post flight ratings - even with the assistance of video recordings - must be less reliable.

The authors are unaware of any other workload rating scale that has been used as extensively for assessing workload in the 'real world'; nevertheless, there are a number of possible shortcomings.

Several authors have underlined the importance of sensitivity and diagnosticity, others have stressed the multidimensional nature of workload, and rating scales having varying degrees of sophistication have been designed with these issues in mind[20-22]. A subjective workload assessment technique (SWAT) developed by Reid and his co-workers[23,24] considers workload in three dimensional terms - time load, mental effort, and psychological stress. Hart and her colleagues[25,26] have developed and refined a multi-dimensional rating scale consisting finally of six subscales, namely: physical demands, mental demands, time pressure, own performance, effort and frustration.

Certainly, in many situations it may be quite important to be able to analyse the reasons why workload has changed. The Bedford scale does not have the power of diagnosis but in practice, on the rare occasions when diagnostic information is required, post-flight discussions with the pilots - especially when their beat-to-beat heart rate plots are used as an aide memoire - are proving to be of considerable value. And, in most cases, assessment of overall or global workload is all that is required - for example, during workload assessments for the purpose of crew complement certification.

The lack of sensitivity at the lower end of the scale was at first thought to be a disadvantage, but experience now suggests that it is unrealistic to strive for a high level of sensitivity. This is particularly so in view of the variations in subjective evaluations between pilots and, occasionally, within the same pilot from time to time. In addition, the cost effectiveness of modifying systems or procedures to correct for small differences in satisfactory workload is questionable.

The nonlinear nature of the scale, although making it difficult to carry out statistical treatments on data, does not seem to cause any problem in practice.

The decision to develop a scale giving relative - rather than absolute - values of workload has not appeared to pose any problem from the practical point of view. And it may be questionable whether absolute rating scales are really necessary. Lidderdale[14], in comparing in-flight with post-flight workload assessments, wrote: "It is possible that all assessments of workload are made from a baseline of comparison with other elements in the flight and, if this is the case, all rating methods may be relative."

Examination of several hundred workload ratings given during a variety of flight tasks or flight phases together with attempts to correlate ratings with heart rate responses have resulted in three important findings. Firstly, as suggested by Ellis[6] in 1979, it was not possible to compare ratings for different flight tasks; for instance, ratings given during the take-off could not be related to those given during the approach and landing.

Secondly, although reasonably consistent, ratings were highly individual for each pilot; unless some normalisation procedure was applied to the data - if sufficient data were available - each subject pilot had to be considered as his own control; a similar but more marked idiosyncrancy has already been identified for heart rate responses[25].

Finally, a good correlation between workload ratings and heart rate responses for the same flight task was apparent in about 80% of pilots but one in five did not show any agreement. The reason for this lack of agreement has not always been identified, sometimes it was because a pilot has not integrated his workload over the entire period of interest, but there is some evidence that it may be related to the nature of that particular individual's heart rate response.

5 CONCLUSIONS

The Bedford workload rating scale was designed for use in the 'real world' of practical flight testing some 10 years ago; and pilots have, without exception, found it easy to use in flight. During the past decade, despite the shortcomings referred to above, its value has been demonstrated in several flight trials - especially when ratings have been augmented by recording the pilots heart rate.

However, unlike the studies carried out on some other rating scales[7,8,9,26,27], the Bedford scale has not been subjected to a critical evaluation in controlled laboratory experiments. Nevertheless, the authors believe the scale in its present form is quite suitable for assessing workload in most practical situations.

REFERENCES

No.	Author	Title, etc
1	R.P. Harper	Pilot evaluation of handling qualities. Ninth Annual Human Factors Society Meeting (1965)
2	J.D. McDonnell	An application of measurement methods to improve the quantitative nature of pilot rating scales. *IEEE Transactions on Man-Machine Systems* **10**, 81-92 (1969)
3	G.E. Cooper R.P. Harper	The use of pilot rating in the evaluation of aircraft handling qualities. NASA Technical Note TN-D-5153 Washington DC (1969)
4	S.J. Gerathewohl	Definition and measurement of perceptual and mental workload in aircrew and operators of Air Force weapon systems. In AGARD CP-181 Higher mental functioning in operational systems AGARD, Paris (1976)
5	G.A. Ellis A.H. Roscoe	The airline pilot's view of flight deck workload: A preliminary study using a questionnaire. RAE Technical Memorandum FS(B) 465 (1982)
6	G.A. Ellis	Subjective assessment, pilot opinion measures. In: AGARDograph No 233 Roscoe A H (Ed) Assessing Pilot Workload, AGARD, Paris (1978)
7	D. McDonnell	Pilot rating techniques for the estimation and evaluation of handling qualities. AFFDL - TR-68-76 Wright-Patterson AFB, Ohio (1968)
8	R.A. Hess	Non adjectival rating scales in human response experiments. *Human Factors*, **15**, 275-280 (1973)
9	D.A. Spyker	Development of techniques for measuring pilot workload. NASA CR-1888, Washington DC (1971)

REFERENCES (continued)

No.	Author	Title, etc
10	J.M. Rolf *et al*	Evaluating measures of workload in a flight simulator. In AGARD CP-146 Simulation and Study of high workload operations. AGARD, Paris (1974)
11	A.H. Roscoe	Assessing pilot workload in flight. In: Conference Proceedings No.373. Flight Test Techniques, AGARD, Paris (1984)
12	A.H. Roscoe	Pilot workload and economic Category 3 landings. In: Conference Pre-prints Aerospace Medical Association Annual Scientific Meeting, Washington DC (1980)
13	W.A. Wa[illegible]right	Flight test evaluation of crew workload. In AGARDograph No 282 Roscoe A H (Ed) The practical assessment of pilot workload, AGARD, Paris (1987)
14	I.G. Lidderdale	Measurement of aircrew workload during low-level flight. In AGARDograph No 282 Roscoe A H (Ed) The practical assessment of pilot workload, AGARD, Paris (1987)
15	T.L. Saaty	The analytical hierarchy process. McGraw-Hill (1980)
16	H.C. Muir R. Elwell	The assessment of workload in helicopters. In: Proceedings of RAE/AGARD/CIT Symposium - The practical assessment of pilot workload, Cranfield (1986)
17	R. Barnes R.C. Graeber	Flight deck environment and working in North Sea helicopter operations. In Conference Abstracts Aerospace Medical Association Annual Scientific Meeting, Washington DC (1988)
18	A.H. Roscoe B.S. Grieve	The impact of new technology on pilot workload. SAE Technical Paper 861773 (1986)

REFERENCES (concluded)

No.	Author	Title, etc
19	A H. Roscoe B.S. Grieve	Assessment of pilot workload during Boeing 767 normal and abnormal operating conditions. SAE Technical Paper 881382 (1988)
20.	J.H. Skipper C.A. Rieger W.W. Wierwille	Evaluation of decision-tree rating scales for mental workload estimation. *Ergonomics*, **29**, 585-599 (1986)
21	G.B. Reid *et al*	Development of multi-dimensional subjective measures of workload. Proceedings of the International Conference on Cybenetics and Society, 403-406, New York (1981)
22	S.G. Hart L.E. Staveland	Development of NASA-TLX (NASA-Task Load Index): Results of empirical and theoretical research In: Human Mental Workload, Eds P.A. Hancock and N. Meshkati, North Holland Press, Amsterdam (1988)
23	G.B. Reid C.A. Shingledecker F.T. Eggemeir	Application of conjoint measurement to workload scale development. In: Proceedings of the Human Factors Society - 25th Annual Meeting (1981)
24	F.V. Schick R.L. Hann	The use of subjective workload assessment techniques in a complex flight task. In AGARDograph No 282 The practical assessment of pilot workload. Ed A.H. Roscoe, AGARD, Paris (1987)
25	M.R. Bortolussi B.H. Kantowitz S.G. Hart	Measuring pilot workload in a motion base trainer. A comparison of four techniques. Proceedings of the Third Biannual Symposium on Aviation Psychology, Columbus, Ohio (1988)
26	W.W. Wierwille J.H. Skipper C.A. Rieger	Decision-tree rating scales for workload estimation: Theme and variations. Proceedings of 20th Annual Conference on Manual Control, Sunnyvale (1984)
27	P. Tsang W.W. Johnson	Cognitive demands in automation Aviat. *Space Environ Med*, **60**, 130-5 (1989)

REPORT DOCUMENTATION PAGE

Overall security classification of this page

UNLIMITED

As far as possible this page should contain only unclassified information. If it is necessary to enter classified information, the box above must be marked to indicate the classification, e.g. Restricted, Confidential or Secret.

1. DRIC Reference (to be added by DRIC)	2. Originator's Reference	3. Agency Reference	4. Report Security Classification/Marking
	RAE TR 90019		UNLIMITED

5. DRIC Code for Originator	6. Originator (Corporate Author) Name and Location
7672000H	Royal Aerospace Establishment, Bedford, UK
5a. Sponsoring Agency's Code	6a. Sponsoring Agency (Contract Authority) Name and Location

7. Title: A subjective rating scale for assessing pilot workload in flight: a decade of practical use

7a. (For Translations) Title in Foreign Language

7b. (For Conference Papers) Title, Place and Date of Conference

8. Author 1. Surname, Initials	9a. Author 2	9b. Authors 3, 4	10. Date	Pages	Refs.
Roscoe, A.H.	Ellis, G.A.		March 1990	16	27

11. Contract Number	12. Period	13. Project	14. Other Reference Nos.
			Flight Management 6

15. Distribution statement

(a) Controlled by –

(b) Special limitations (if any) –

If it is intended that a copy of this document shall be released overseas refer to RAE Leaflet No.3 to Supplement 6 of MOD Manual 4.

16. Descriptors (Keywords) (Descriptors marked * are selected from TEST)

Workload. Flight. Pilot. Rating scale.

17. Abstract

Despite the many techniques developed for evaluating pilot workload in flight subjective assessment by experienced pilots is still the most reliable method by far. This report describes the design and development - with the help of practising test pilots - of a ten-point rating scale. The scale uses a decision tree similar to that used by the Cooper-Harper Handling Qualities scale, and is based on the concept of spare capacity. Examples are given of its use by a large number of pilots in various flight trials and workload studies.

FS9(b)