EXPERIMENTAL ASSESSMENT AND ANALYSIS OF THE CREW-WORKLOAD

Prof. Giuseppe Surace – Polytechnic of Turin – Aerospace Department
Juri Barollo – Alenia Aeronautica S.p.A. – Systems and Simulation Laboratories

Keywords: *Workload, Human Information Processing, Human Machine Interface*

Abstract

The overcrowded air traffic scenario and the steady hi-tech growth within the cockpits of new aircraft may impose levels of workload on aircrew that deserve attention.

The extended duty time and the mental workload due to the new monitoring tasks introduced by the "glass cockpit" increase the possibility of lack of situational awareness.

The topic of workload assessment is hence an interesting one for flight safety.

In collaboration with AirOne, a survey of the overall workload felt by aircrew flying with B737 during short and medium range flight has been carried out using the NASA TLX (Task Load Index) methodology.

A brief discussion of workload concept and workload methodologies assessment are given and a survey conclusion is reported.

1. The new role of the aircrews

A lot of aviation incidents/accidents are related to the so called "Human Factors"; but how can we define it?

Despite the term Human Factors is coming to acquire a wider meaning, we can say that Human Factors is about people in their working and living environments. Human Factors is about the relationship between people and machines, equipment, procedures and with the environment about them. From another point of view, Human Factor can also be seen as a technology devoted to the effectiveness of the systems, which includes safety and efficiency, and well-being of the individual [1].

The Human Factor considerations during the design of new human-machine systems are addressed to the development of working stations and working environments in which the possibility of human error is minimized. This is particularly true in the aeronautical context. In the last decades, we have seen radical changes in aircraft systems and cockpit environment. The advance in electronic and informatics technologies has led to a high level of automation in flight controls and flight management. The development of CRT (Cathode Ray Tube) displays and, more recently, LCD (Liquid Crystal Display) flat panels have radically changed the quantity and the format of the information displayed to aircrews.

Aircrews flying modern hi-tech aircraft have to face a new role in the aircraft control loop, a role of manager rather than of bold pilots! Better qualities of "airmanship" are now required as well as high levels of "cognitive performance" and "decision making".

2. Human Processing Limitations

Dedicated pilot training sessions are intended to develop in the aircrew an adequate level of airmanship (i.e. Crew Resources Management and Multi Crew Coordination courses), so that, if we neglect some crucial human inherent and operational factors we could affirm that a trained aircrew could ensure a safe flight.

Unfortunately we cannot.

On 8 May 1978, National Airlines Flight 193, a Boeing 727, crashed into Escambia bay while on a surveillance-radar approach to Pensacola Regional Airport. The NTSB (National Transportation Safety Board) determined that the probable cause of this accident was the crew's unprofessionally conducted non-precision instrument approach in which the pilots failed to monitor the descent rate and altitude. Contributing to the accident was the radar controller's failure to provide advance notice of the start descent point, which accelerated the pace of the crew's cockpit activities after the passage of the final approach fix.

This created confusion in the cockpit at which a series of mistakes followed including failure to extent the gear and flaps appropriately and failure to heed a warning from the ground proximity warning system. The increased aircrew's workload created a situation that would make impossible for the captain to configure his aircraft in the manner specified in the flight manual and contributed to a lack of attention to vital information concerning the aircraft's altitude, speed, and descent rate [2].

This tragic example show that, despite training and experience, the inherent human limitations in information processing can lead to the human error.

Everyone can test for themselves the limits on the number of activities one can handle concurrently. Dedicated studies on the topic show that the total amount of activities one can deal with results not only from the inherent characteristics of the brain and nervous systems but also from the nature of activities involved and the extent to which they draw on time-shared common resources.

The ability of pilots to perform multiple tasks has been shown to be related to the processing demands imposed by the individual tasks. Monitoring the PFD (Primary Flight Display) can be thought of as a visual and spatial task, whereas listening an ATC (Air Traffic Control) communication is essentially an aural and verbal task. Since has been shown by physiological and psychological research that spatial and verbal processing operate in quite different ways and even take place in separate parts of the brain, it follows that there is less conflict between these two tasks than, for example monitoring the PFD and copying a departure clearance.

This is in accordance with the multiple-resource model proposed by C.D.Wickens for human information processing [3].

This model suggests that the limits of information processing depends on the extent of residual attentional resources that can be switched from one area to another of the information processing pattern during contemporaneous tasks.

The application of this principle has led to improvement in cockpit human-machine interface design. Two example are voice-activated control systems and auditory displays. Such systems and displays are less likely to interfere with the primary task of flying than others that compete for the same resources.

3. Workload and fatigue

As stated above, the modern aircrews have to manage several sources of information in order to achieve and maintain a continuously updated *mental model* of the flight in progress. Only doing so the aircrew will be able "to fly ahead the airplane".

Such management activity must be performed assuring a stated level of performance during each phase of flight, facing contemporaneous tasks as flying the airplane, communications, etc.

The notion of a trade-off in performing contemporaneous tasks has formed the basis of most attempts to define and measure cockpit workload.

Workload is a general term used to describe the cost of accomplishing task requirements for the human element in the man-machine system; this cost may be reflected in depletion of attentional, cognitive or response resources, inability to accomplish additional activities, emotional stress, or performance deficits.

Although an official definition of workload still does not exist, workload can be described generally in terms of the relationship between the demands imposed upon the operator by the task, and the capacity of the operator to meet those demands.

A number of proposed definitions of workload are:

- "workload represents the costs incurred by the human operator in achieving a particular level of performance" [4];

- "workload is a hypothetical construct that reflects the interaction between a specific individual and the demands imposed by a particular task" [5];
- "workload is the effort invested by the human operator into task performance" [3];

Definitions are not correct or incorrect but rather more or less useful.

It must be specified that concept of workload includes mental and physical workload. Physical workload is a straightforward concept; it is easy to define and measure in terms of energy expenditure. Traditional Human Factors texts tell us how to measure human physical work in kilocalories and oxygen consumption.

Measures of physical workload are becoming less and less relevant in aviation where hydraulic systems and other devices have been allocated the function of exerting large forces that once were the responsibility of the human operator; this has meant that the physical component of workload is becoming less and less relevant than mental workload.

Only one aspect related to the physical feeling still remains a topic of interest: fatigue and its management.

Workload and fatigue are correlated but different concepts, influencing each other.

As for workload, the definition of fatigue is a source of difficulty which may generate confusion. Four interpretations of fatigue may be considered. Firstly, it may reflect inadequate rest; secondly it may refer to symptoms associated with disturbed biological rhythms described often as the "jet lag" effect. Thirdly, it may be due to excessive muscular or physical activity and, finally, fatigue could result when excessive cognitive work has been undertaken.

If we consider a typical working day of a typical aircrew (with its duty time that often exceed the ten hours due to pre-flight preparation, crew check-in time, passengers check-in, customs and immigration formalities at the end of flight, etc.) and the cognitive tasks imposed by the modern airplane, we can sense the thin bond between fatigue and workload.

4. Workload drivers

In some phases of flight (i.e. the approach) and in particular environmental conditions, situations may occurs in which the information rate that crew must perceive and analyse could exceed their cognitive processing capability, increasing the likelihood of human error.

Also during low levels of task demands human error may lie in wait. Boring monitoring time during the cruise phase of flight (characterised by low workload) can cause a low level of arousal. Falling into microsleeps is common during trans-oceanic flight.

Although aircrew workload and performance are related, the nature of this relationship is not straightforward. The literature shows the "inverted U-shape relationship" (fig.1) between level of arousal/workload and task performance.

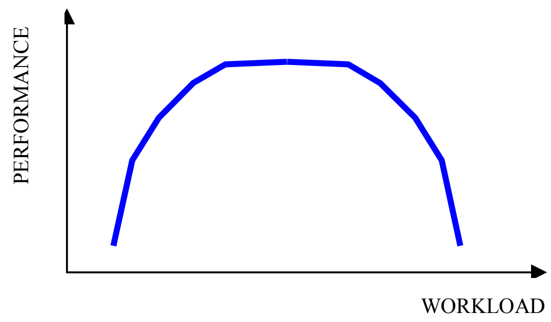


Fig.1: performance versus workload relationship

Such a qualitative function shows that the optimum level in task performance is achieved only with an optimum level of workload.

During the design phase of human-machine systems the maximum effort must be made in the application of human-machine theory in order to meet such an optimum trade-off.

The workload assessment, therefore, must be considered also as a design tool for new human-machine systems or improvement of already existing ones.

Is it possible define particular workload drivers?

Trying to give an answer to the question in the aeronautical context, some general workload drivers have been identified: time pressure, flying qualities (control dynamics), working memory load and meteorological conditions.

Time pressure is the dangerous condition that arises when the time required for optimal task execution is less than the time available. It may be due to external factors (such as unexpected task) or internal factors (such as poor task management policy). Since there are substantial individual differences in the effectiveness of tasks/workload management, the aircrews should be trained to identify optimal task hierarchies during the MCC-CRM training courses.

The aircraft *flying qualities* play a non negligible role in the amount of spare capacity and resources that pilot can allocate to other tasks. Flying with a "nervous airplane" could make it impossible to perform cognitive tasks!

The *working memory load* that crew experience to perceive, analyse and manage the flow of available information is maybe the most relevant workload driver in the modern cockpit environment. Data presented by displays that are not understood or that "apparently" do not match with other available information, may result in an overload of the pilot's cognitive process. Only a careful design of human-machine interfaces and procedures, with proper training, can avoid these dangerous situations.

Aircraft operate in a extremely variable atmospheric environment, hence, we should not forget *meteorological conditions* as a workload driver. The effect of adverse meteorological conditions is essentially to increase the task demands due to the further stress that pilots have to face (low visibility, icing conditions, turbulence, etc.)

5. Workload measurement methodologies

Workload is a complex, multidimensional phenomenon. It is difficult to define precisely, but rather easy to recognise when it becomes uncomfortably high! It would be helpful to have objective measures of workload, but its cognitive character, behavioural and physiological variables and practical constraints do not make workload measurement a simple matter.

The different methodologies used to measure aircrew workload tend to fall into four categories: subjective ratings, primary task measures, secondary task measures and physiological measures.

Such different kinds of measures differ by their sensitivity, diagnosticity, intrusiveness, implementation requirements and pilot acceptance.

Sensitivity refers to the capability of a technique to detect changes in the levels of workload imposed by task performance.

Diagnosticity refers to the capability of a technique to indicate not only when the level of workload varies, but also the cause of such variation; it should indicate which of the capacities or resources are varied by demand changes in the system.

Intrusiveness refers to the amount of contamination exerted by the workload measure on the operator primary task performance; the intrusiveness level has to be carefully evaluated if workload is being assessed while the crew is performing its duty in the real operational environment in order not to jeopardize flight safety.

Implementation requirements refers to all the practical constraints dealing with the complexity of the measurement procedures and apparatus. This includes things like the instrumentation and software necessary for data collection and the level of operator training required before valid results can be obtained.

Pilot acceptance: the choice of proper workload measurement techniques should also be evaluated with respect to the pilot's perception of the validity and utility of the technique itself; workload assessment procedures perceived as intrusive or "artificial" involve the risk of being ignored or performed at substandard levels, thus compromising potential effectiveness. So, care must be taken to explain to the crew the purpose of workload assessment techniques and associated measurement procedures.

It's really difficult to arrive at a single unitary measure of workload since the different kinds of methodology have their advantages and disadvantages. They meet the criteria stated above in different ways.

Subjective ratings performed by the aircrew or an expert observer may come closest to tapping the essence of workload. They provide a general and sensitive measure of workload. Furthermore, since they can be performed at the end of each flight, they do not intrude or affect the primary task of flying.

Physiological measures (such as pulse rate, heart rate variability, eye blink rate, etc.) seem to have often low levels of sensitivity, diagnosticity and pilot acceptance with respect to the low levels of intrusiveness due to instrumentation and equipment needed. At the other hand, they reflect the overall activation level of the subjects.

Primary and secondary task measures are related to the notion of spare processing capacity. While the pilot is performing a primary task (such as, for example, flying a flight profile with stated heading, altitude and velocity), he is also required to perform a secondary concurrent task such as a mental arithmetic task, a communication task, an estimation of elapsed time and so on. Secondary tasks should not intrude on the primary task for competing demands on motor channels as well as central resources and it should also preserve the priority of the primary task. This is essential if the test is carried out in the real operational environment.

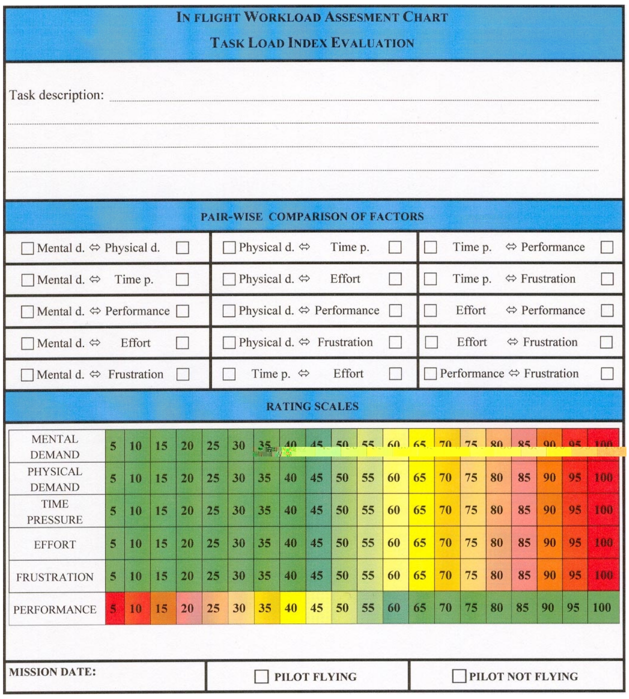
The decay in both primary and secondary task performance is indicative of the spare processing capacity.

Researchers agree that, attempting to assess a complete workload profile, it would be good practice to use, if possible, a multiple set of measures.

6. The application of a subjective rating scale

Attempting to evaluate the level of workload that civil transport aviation aircrews have to face in the real operational environment, the application of a non intrusive technique as a subjective rating has been performed thanks to the kind collaboration of AIRONE, an Italian civil transport company.

Pilots of AIRONE have been requested to fill in the workload data sheet shown in fig.2. This has been inspired by the NASA-TLX (Task Load Index) methodology developed by Hart and Steveland [6] at the NASA Ames Research Center.



IN FLIGHT WORKLOAD ASSESMENT CHART

TASK LOAD INDEX EVALUATION

Task description: ....

PAIR-WISE COMPARISON OF FACTORS

☐ Mental d. ⇔ Physical d. ☐	☐ Physical d. ⇔ Time p. ☐	☐ Time p. ⇔ Performance ☐
☐ Mental d. ⇔ Time p. ☐	☐ Physical d. ⇔ Effort ☐	☐ Time p. ⇔ Frustration ☐
☐ Mental d. ⇔ Performance ☐	☐ Physical d. ⇔ Performance ☐	☐ Effort ⇔ Performance ☐
☐ Mental d. ⇔ Effort ☐	☐ Physical d. ⇔ Frustration ☐	☐ Effort ⇔ Frustration ☐
☐ Mental d. ⇔ Frustration ☐	☐ Time p. ⇔ Effort ☐	☐ Performance ⇔ Frustration ☐

RATING SCALES

MENTAL DEMAND	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
PHYSICAL DEMAND	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
TIME PRESSURE	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
EFFORT	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
FRUSTRATION	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
PERFORMANCE	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100

MISSION DATE:	☐ PILOT FLYING	☐ PILOT NOT FLYING

Fig.2: Workload data sheet inspired to NASA-TLX methodology

This methodology, in spite of the limitations that could arise from the application of only one technique, is able to provide an overall aircrew workload index and also, thanks to its multidimensional character, is able to indicate which are the relevant sources of it.

The overall workload score is based on a weighted average of magnitude ratings on six subscales: Mental demands, Time pressure, Performance, Effort and Frustration. The importance of each factor as a source of workload for a particular task is obtained by a simple pair-wise comparison among the six factors.

Ratings on each subscale are obtained after each performance of the task. By giving more weight to ratings of factors that are most important during a particular task, the sensitivity and diagnosticity of the derived workload score is enhanced.

There are 15 possible pair-wise combinations of the 6 scales. The number of times each factor is selected as being more relevant to the workload of a particular task, in comparison to each other factor, is tallied.

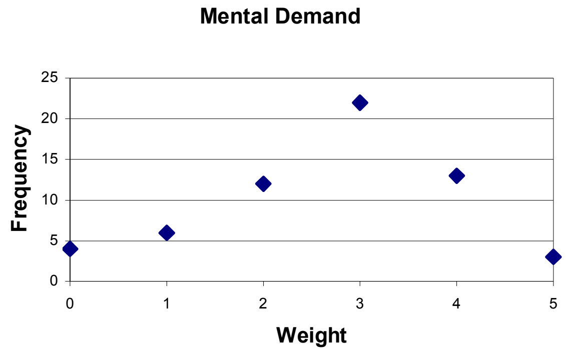
The minimum for each factor is 0 (not at all relevant), the maximum tally is 5 (more important than other factor).

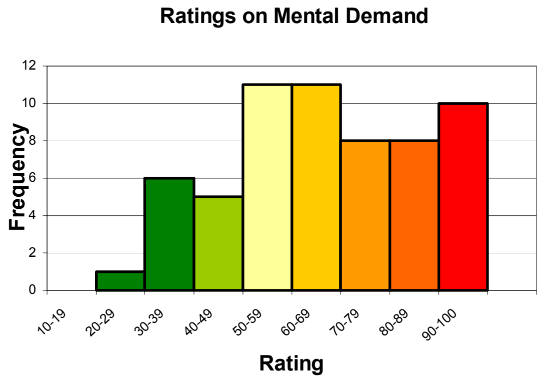
Ratings are then given individually on each factor; each scale is divided into 20 equal intervals from 0 to 100 in increments of 5 points and painted with an intuitive shaded colour code (from green for “good” to yellow for “caution ” and red for “dangerous”).

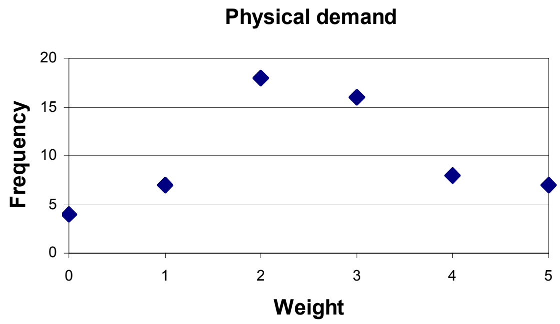
The weights and the ratings may or not may covary. For example, it’s possible for mental demand to be the primary source of loading for a task, but the magnitude of those demands might be low.

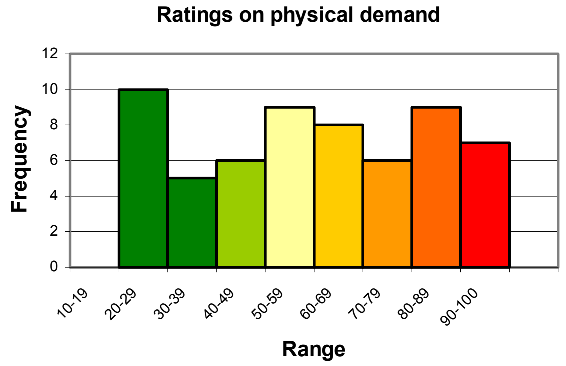
AIRONE’s aircrew have been trained in the meaning of single dimensions to be rated and also in the way to complete the data sheet. The pilots have also been requested to provide pair-wise and ratings with regard to the overall flight during post-flight debriefing. This highlights the unintrusiveness of the technique and the practical absence of implementation requirements (a pen and a sheet are enough!).

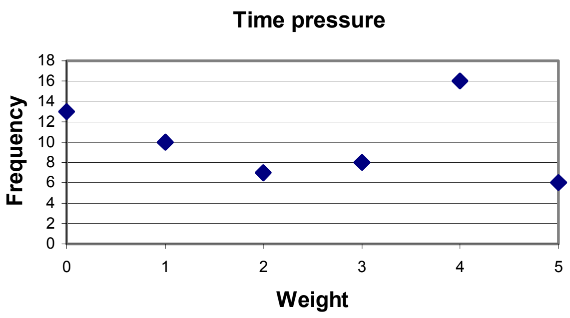
A set of sixty flights flown in different meteorological conditions has been examined. In the following graphs the data collected are shown as frequency of weight and ratings for each dimension.

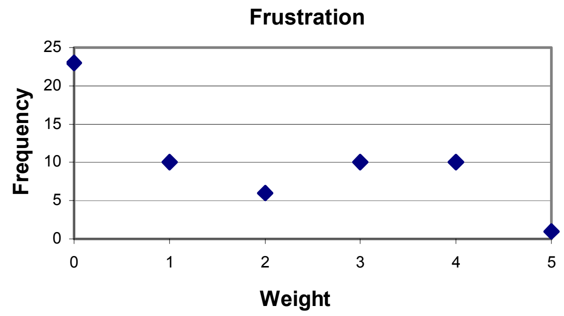


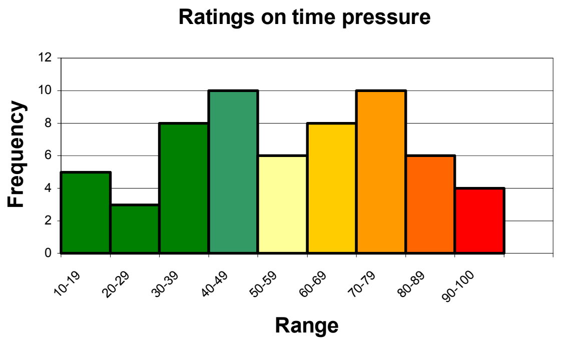


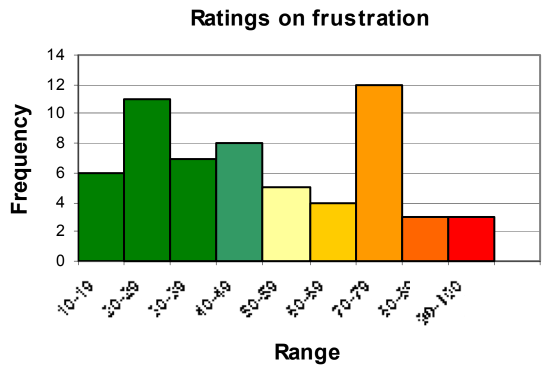


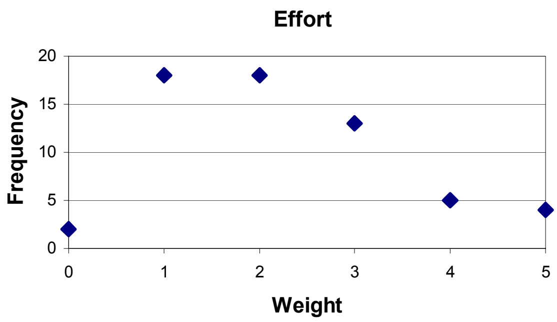


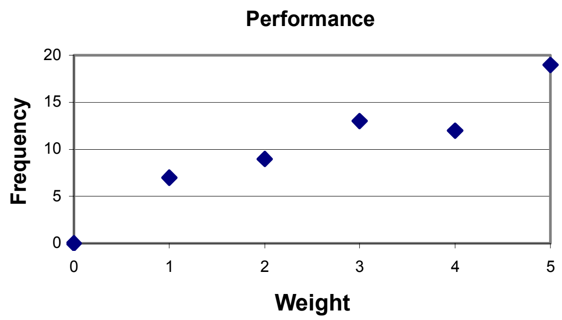

Time pressure
Frequency
Weight

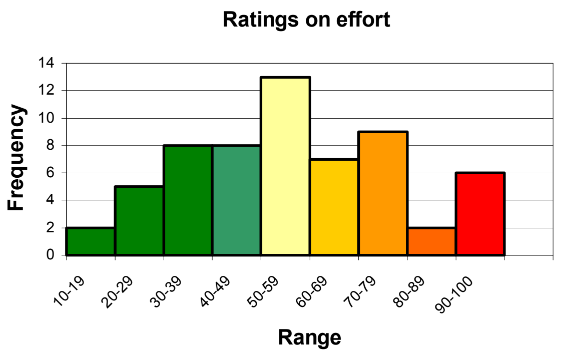

Frustration
Frequency
Weight

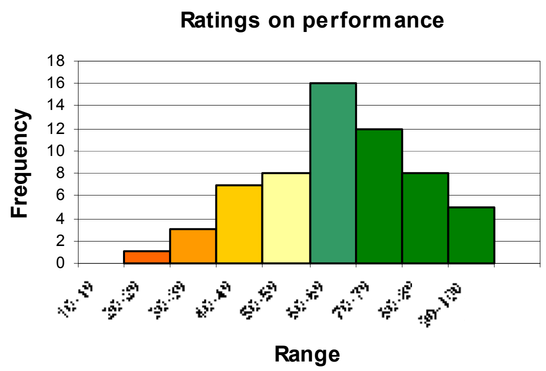

Ratings on time pressure
Frequency
10-19
20-29
30-39
40-49
50-59
60-69
70-79
80-89
90-100
Range

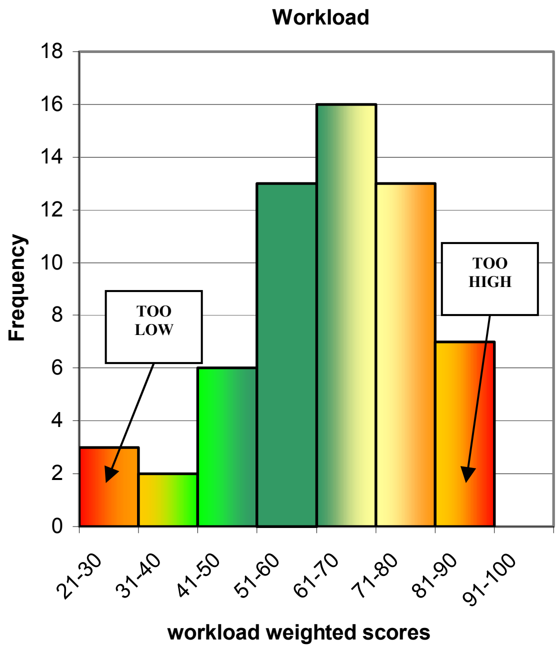

Ratings on frustration
Frequency
Range


Effort
Frequency
Weight


Performance
Frequency
Weight


Ratings on effort
Frequency
10-19
20-29
30-39
40-49
50-59
60-69
70-79
80-89
90-100
Range


Ratings on performance
Frequency
Range



7. Conclusions

Despite the survey sample not being really large, giving a quick glance to the graphs some preliminary conclusions may be drawn.

First of all, this little survey seems to confirm cognitive load as a relevant source of workload for the aircrew. Despite the mental demand dimension not being frequently felt by the aircrew, its ratings are quite indicative of a considerable cognitive resource expenditure.

Physical demand shows an unexpected rating level . To explain this, it is worth considering at the same time the effort and frustration dimensions. In fact, with this simultaneous view it is possible to have a look in the dimension of "fatigue" felt by the crew. We note that effort, despite with not relevant weight, is frequently felt. The frustration dimension graph shows situations of considerable rating values. Such integration between the two previous dimensions may reflect a status of fatigue that could increase the perceived level of physical demand imposed by a typical flight profile.

The time pressure dimension is quite distributed both in terms of weight and ratings. This is probably due to the application of multi-task management techniques learnt by the crew during the MCC-CRM training courses; there still remain, of course, situations of relevant time pressure due to, probably, unexpected operational delays (i.e. slow aerodrome procedures) or unexpected variations in the planned flight (i.e. variations in the approach procedures, etc.).

The Performance dimension is frequently highly weighted but not often highly rated. It seems that aircrew try to get the best performance but it's often not able to achieve it! Therefore the frustration may be increased.

The overall workload score trend reflects the results achieved with the application of human engineering concepts to the design of complex systems such a cockpit environment. The most frequent values of overall workload fall, in fact, in the middle between the too high and too low regions.

This survey must be interpreted as a look into the fascinating world of Human Factors in aviation research. An improvement would be to analyse of different phases of a typical flight (i.e. taxi, take-off, climb, cruise, approach and landing) with regard to the different crew positions (i.e. pilot in command and co-pilot).

Acknowledgements

The authors appreciate the kindly collaboration of Cap. Giorgio Amoroso (AIRONE Flight Safety Director); special thanks to all the aircrews of AIRONE for their collaboration.
This paper has been written with the patient collaboration of Sara Fantino, to whom the paper is dedicated.

References

[1] Frank H. Hawkins. *Human* 257.8 *in Flight*. Orlady Edition

[2] O'Hare and Roscoe. *Flightdeck Performance*. Iowa State University Press.

[3] Wickens C.D. *Attention, Time Sharing and Workload* in Engineering Psychology and Human Performance. Merril Edition

[4] Muir Hellen. *Introduction to Human Factors in aviation: pilot workload assessment*.3/5 May 2000, Cranfield University.

[5] Gopher and Donchin. *Workload: an examination of the concept* in Handbook of perception and performance. Wiley Edition, New York (1986)

[6] Hart and Steveland. *Development of NASA TLX: results of experimental and theoretical research* in Human Mental Workload. Hancock Edition.