Workload Functions Distribution Method: A Workload Measurement Based on Pilot's Behaviors

Yiyuan Zheng[1(✉)], Yuwen Jie[1], Tong Zhang[1], and Shan Fu[2]

[1] Shanghai Aircraft Airworthiness Certification Center of CAAC, Shanghai, People's Republic of China
leodeisler@sjtu.edu.cn
[2] Shanghai Jiao Tong University, Shanghai, China

Abstract. According to the airworthiness regulation, FAR25.1523 published by Federal Aviation Administration, the minimum flight crew must be established so that it is sufficient for safe operation considering the workload on individual crew members. Considering workload evaluation, typically, the measurements classified into three types: performance measures, subjective rating scale measures and psychophysiological measures. However, although these measurements are widely used in various fields, they could not reflect the behavior of flight crew during flight tasks comprehensively, especially their workload functions. Normally, the basic workload functions based on the flight crew behaviors consist of six aspects: flight path control, collision avoidance, navigation, communication, operation and monitoring of aircraft engines and systems and command decisions. In this study, upon the above six aspects, a measurement named Workload Function Distribution Method was developed, considering flight crew behaviors including fixations, actions and communications in flight tasks. In order to verify the Workload Function Distribution Method, three flight tasks with different complexity were carried out among 6 flight crews in a CRJ200 flight simulator. The three flight tasks were standard instrument approach, non-precision approach with normal weather condition and non-precision approach with turbulence. Furthermore, one of the subjective rating scale measures, NASA-TLX, was used as a verification method which collected after each task. The experiment results indicated that Workload Function Distribution Method could distinguish the different complexity of flight tasks, and related to the NASA-TLX Scale.

In conclude, Workload Function Distribution Method was built up in this study. This measurement could effectively represent the flight crew workload based on their behaviors in flight tasks.

Keywords: Fixation · Operation · Workload · Workload function

1 Introduction

As described in the airworthiness regulation [1] and the related Mean of Compliances [2], the minimum flight crew was established by considering the workload on individual crew members. However, how to define workload is still controversial. Several

D. Harris (Ed.): EPCE 2016, LNAI 9736, pp. 48–55, 2016.
DOI: 10.1007/978-3-319-40030-3_6

researchers have given out some acceptable definitions. Stassen et al. regarded the workload as the mental effort that human operator devotes to control or supervision relative to his capacity [3]. Eggemeier et al. assumed that the mental workload refers to the portion of operator information processing capacity or resources that is actually required to meet system demands [4]. Kramer and Sirevaag held that the workload was the cost of performing a task in terms of a reduction in the capacity to perform additional tasks that use the same processing resource [5]. According to these definitions, the workload should be considered as a capacity proportion rather than an exact value. Typically, workload measurement could be divided into three types in aviation: performance measures, subjective rating scale measures and psychophysiological measures [6]. In aviation, performance measures include examining the deviation between actual flight and expected path [7], and response time and accuracy of flight crew [8]. The most widely used subjective rating scale measures are Modified Cooper-Harper scale, NASA-TLX, and Bedford Scale [9]. Psychophysiological measures usually considered eye movement, heart rate and respiration [10]. However, all these measures have their own limitations. Wickens noted that performance measures may not be sufficient or adequate in some conditions [11]. Subjective rating scale measures are sometimes uncertain on the repeatability and validity, and data manipulations are often questioned as being inappropriate [12]. Psychophysiological measures are influenced by other factors, such as heart rate is quite different in day and night [13].

In order to reflect flight crew's workload more specifically and comprehensively, Workload Function Distribution Method, which combined basic workload functions in flight, was built up in this paper. According to CS 25.1523, the basic workload functions depending on the flight crew behaviors consist of six aspects: flight path control, collision avoidance, navigation, communication, operation and monitoring of aircraft engines and systems and command decisions. Workload Function Distribution Method was implemented in three approach scenarios: Standard Instrument Approach, Non-Precision Approach, and Non-Precision Approach with turbulence. Meanwhile, NASA-TLX was used as a comparative and verification indication.

2 Method

2.1 Subjects

Twelve commercial airline male pilots were invited to participate in this experiment. All of them are Chinese. The age range of these pilots was from 30 to 50 years old (Mean = 38.3 ± 7.55yr). Mean total flight hours of them were 6543.6 ± 4680.9 h (range from 1000 to 18000 h), and mean flight hours in the last two weeks before the experiment were 10.55 ± 9.32 h (range from 0 to 80 h). Each pilot has been either captain or co-captain of CRJ-200 for more than 1 year (Mean = 4.32 ± 3.78 yr). Simultaneously, they have all been recruited as captains or co-captains for other types of aircrafts (6 for B737, 4 for A320, and 2 for B747). Before the experiment, all the subjects signed an informed consent form prior to participation.

2.2 Equipment

The experiment was carried out in a CRJ-200 full - flight simulator. It is a qualified flight simulator (level C) conforming to the guidance presented in Federal Aviation Administration Advisory Circular (AC 120-40B) - Airplane Simulator Qualification [14].

Besides the flight simulator, an eye tracker (Tobbi Glass, Sweden), which sample rate was 50 Hz, was used to determine the fixation areas and corresponding durations of the subjects during the experiment. The minimum fixation period was 200 ms.

The performance of each pilot was recorded by a wide angle video camera (pixel: 640*480, sampling frequency: 10 Hz) which was installed behind the emergency exit door on the flight deck. The recording accurately described the actions of pilots over time and space during the experiment.

2.3 Procedure

In order to distribute the subject's behaviors into the six workload functions aspects, several rules were established as following first:

(a) Flight path control

If the operations were carried out on Mode Control Panel (MCP) and Control Wheel, including moving time (i.e., hand moving from one device to another).

If the fixations were focused on Primary Flight Display (PFD).

(b) Collision avoidance

If Traffic Collision Avoidance System (TCAS) alert appeared.

(c) Navigation

If the fixations were focused on Out of Window (OTW) or Multi-Function Display (MFD).

(d) Communication

If the communications existed between flight crews or between crew and Air Traffic Controller (ATC).

(e) Operation and Monitoring of aircraft engines and systems

If the fixations were focused on Engine Indication and Crew Alerting System (EICAS).

If the operations were not carried out on Control Wheel.

(f) Command decisions

If the operations were not dependent on checklist.

The twelve pilots constituted six flight crews, and participated in the experiment including three approach scenarios, which were Standard Instrument Approach, Non-Precision Approach, and Non-Precision Approach with turbulence. Before the experiment, each subject were required to be trained in the simulator for one hour to be familiar with the configurations and the procedures of the scenarios. In the experiment, the subjects conducted the three tasks sequentially. Simultaneously, one flight instructor stayed with the flight crew in the simulator, who was responsible for the configurations of each scenario, and acted as the role of ATC. After each scenario, every subject was asked to fulfill the NASA-TLX scale. The whole duration of the experiment had been last for two weeks.

3 Results

The results of the experiment were divided into five aspects. Firstly, the NASA-TLX scales results were given out to represent the workload differences among the three types of approach. Then, the one function conditions, the two functions conditions and the three functions conditions were analyzed sequentially. Last but not least, the four to six functions were combined together for analyzing. Only the behaviors of Pilots Flying (PF) were provided here.

(a) NASA-TLX scales

From the results of the NASA-TLX scales, the subjective workloads were significantly different among the three types of approach ($F(2, 15) = 3.865$, $p = 0.044$). However, the differences between each two of the three were insignificant. Among them, Standard Instrument Approach had the maximum mean workload (Mean = 48.64, SD = 14.33), Non-Precision Approach had the minimum mean workload (Mean = 31.81, SD = 4.80), and the mean workload of Non-Precision Approach with turbulence was in the middle (Mean = 42.75, SD = 10.54).

(b) One Workload Function

The mean proportions of one workload function comparing with overall task time among the three types of approach was shown as Fig. 1. The difference of one

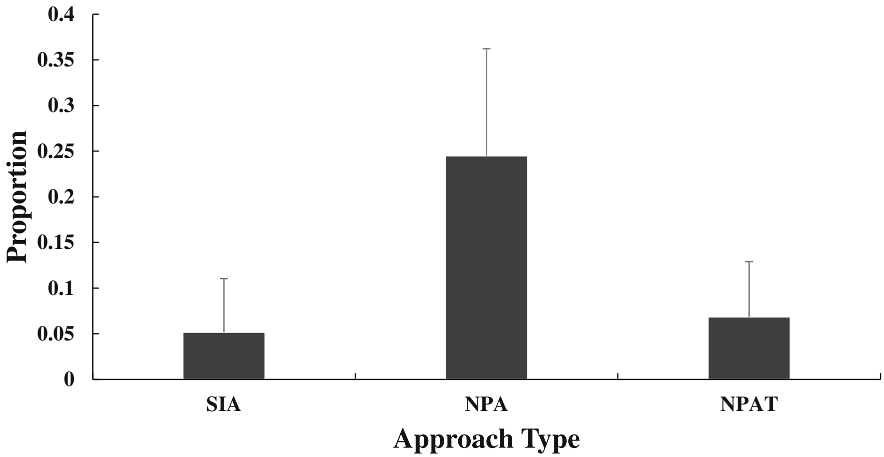

Fig. 1. The mean proportions of one workload function

workload function among the three types was significant ($F(2,15) = 9.903$, $p = 0.002$). An extra T-test was conducted. The differences between Standard Instrument Approach and Non-Precision Approach ($t = 3.614$, $p = 0.005$), and between Non-Precision Approach with turbulence and Non-Precision Approach ($t = 3.280$, $p = 0.008$) were significant. However, the difference between Standard Instrument Approach and Non-Precision Approach with turbulence was insignificant ($t = 0.483$, $p = 0.640$).

(c) Two Workload Functions

The mean proportions of two workload functions comparing with overall task time among the three types of approach was shown as Fig. 2. The difference of two workload functions among the three types was significant ($F(2,15) = 28.798$, $p = 0.000$). An extra T-test was conducted. All the differences between Standard Instrument Approach and Non-Precision Approach ($t = 7.145$, $p = 0.000$), between Standard Instrument Approach and Non-Precision Approach with turbulence ($t = 4.256$, $p = 0.002$), and between Non-Precision Approach with turbulence and Non-Precision Approach ($t = 3.743$, $p = 0.004$) were significant.

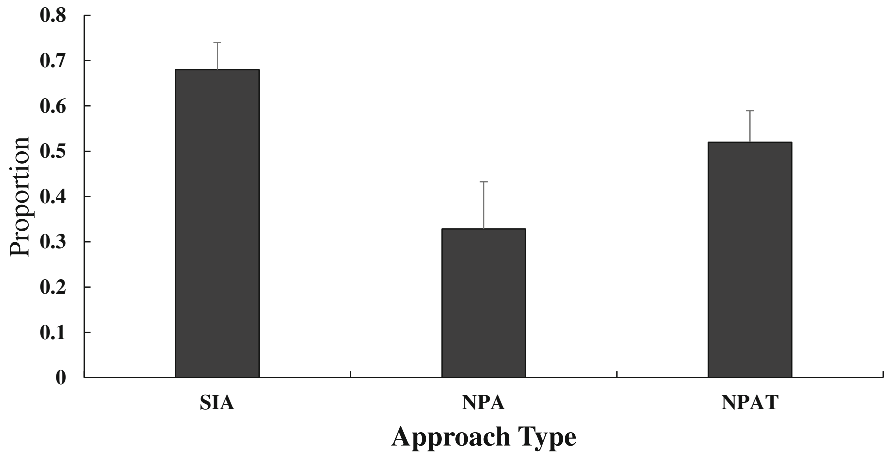

Fig. 2. The mean proportions of two workload functions

(d) Three Workload Functions

The mean proportions of three workload functions comparing with overall task time among the three types of approach was shown as Fig. 3. The difference of three workload functions among the three types was significant ($F(2,15) = 4.344$, $p = 0.032$). An extra T-test was conducted. The differences between Standard Instrument Approach and Non-Precision Approach ($t = 3.502$, $p = 0.006$), and between Non-Precision Approach with turbulence and Non-Precision Approach ($t = 2.486$, $p = 0.032$) were significant. However, the difference between Standard Instrument Approach and Non-Precision Approach with turbulence is insignificant ($t = 0.216$, $p = 0.833$).

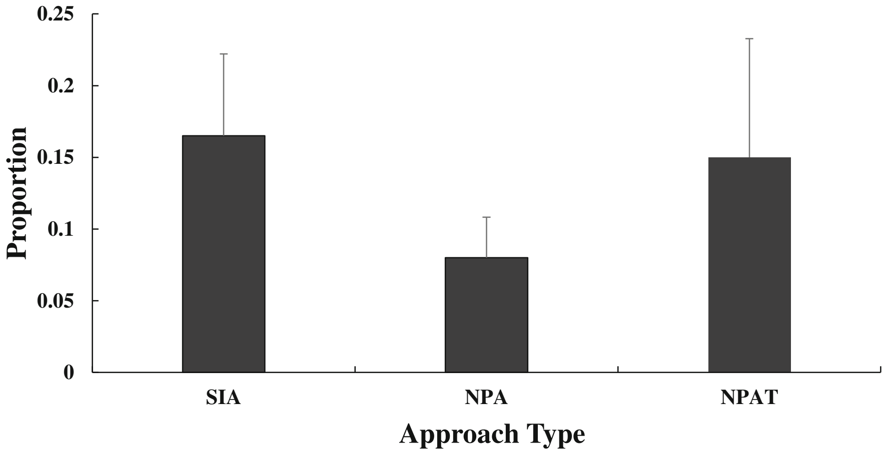

Fig. 3. The mean proportions of three workload functions

(e) Four to Six Workload Functions

The mean proportions of four to six workload functions comparing with overall task time among the three types of approach was shown as Fig. 4. The four to six workload functions difference among the three types was insignificant ($F(2,15) = 1.075$, $p = 0.366$). An extra T-test was conducted. None of differences between Standard Instrument Approach and Non-Precision Approach ($t = 1.046$, $p = 0.320$), between Standard Instrument Approach and Non-Precision Approach with turbulence ($t = 0.157$, $p = 0.878$), and between Non-Precision Approach with turbulence and Non-Precision Approach ($t = 1.763$, $p = 0.108$) were significant.

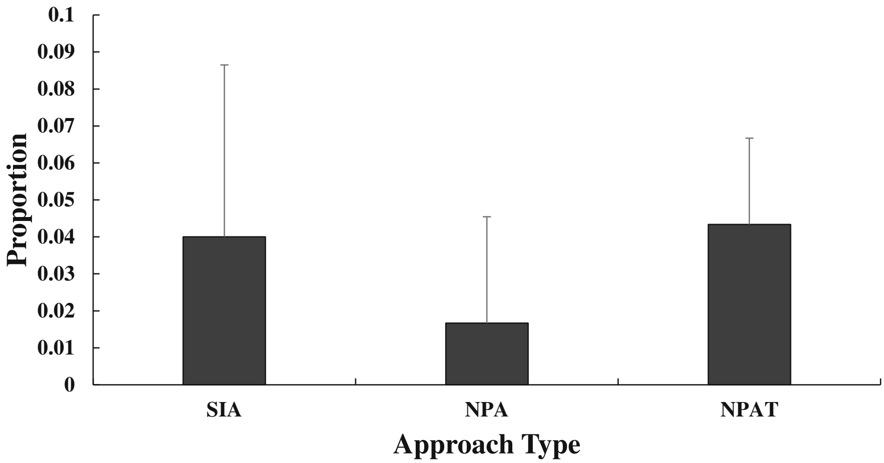

Fig. 4. The mean proportions of four to six workload functions

4 Discussion

Since the NASA-TLX scale is a widely used subjective workload evaluation method, the results of the NASA-TLX scales of the experiment was used as a baseline workload comparison of the three types of approach. The results showed that the three types of approach were significantly different. Besides, all the tasks were acceptable.

Considering the results of the workload functions distribution method, one workload function, two workload functions and three workload functions all had the similar results as the NASA-TLX scales. They all reflected the significant difference among the three types of approach. This could be explained according to Multiple Resource Theory [15]. In one workload function, the subjects usually either observed the flight deck display or carried out an action on control device (i.e., one single channel was actuated in human information processing). In two workload functions, two different channels were occupied simultaneously (i.e., the subjects observed the display and carried out actions concurrently). Time sharing between two tasks was more efficient if the two used different structures than if they used common structures [16]. Buhusi and Mack suggested that in time sharing tasks, brain circuits could reallocate the attentional and memory resources [17]. Similar results were found in three workload functions, in which auditory, visual and manual channels worked at the same time. All these workload functions conditions were tolerable by the pilots, and the different proportions of the workload functions could represent the workload differences on the pilots of the three types of approach. In four to six workload functions, the proportions were less than 0.05, and the difference was insignificant. This was reasonable, because in designing operating procedures of the tasks, to perform over four workload functions simultaneously was unacceptable. Information resource channel conflicts would lead to potential human error [18].

According to the experiment results, Workload Function Distribution Method could efficiently indicate different workload of the tasks, especially in one workload function, two workload functions and three workload functions. Comparing with the tradition three types of workload measurements, this new developed method was directly related to the flight task, and more comprehensive. Furthermore, unlike other measurements, it was scarcely influenced by the ambient environment. Besides, the area of interests (AOIs) was usually used to study the attention allocation or scan pattern of flight crews [19, 20]. Workload Function Distribution Method involved AOIs in evaluating pilot's workload directly.

However, some improvements still should be carried out about Workload Function Distribution Method. Firstly, the usability of the new method is more complicated than other measurements, especially on determining AOIs and the actions of the pilots. Secondly, more tasks with various distinguishable complexity need to validate the method.

5 Conclusion

In this study, a new method named Workload Function Distribution Method was developed based on the pilot's behaviors. Three types of approach were selected as experiment scenarios. According to the results, one workload function, two workload functions and three workload functions all had the similar results as the NASA-TLX scales, and could efficiently indicate pilot's workload. Further work would be carried out to make the method more convenient.

References

1. EASA, Certification Specifications for Large Aeroplanes CS-25 (2009)
2. EASA, CS-25 BOOK2 Acceptable Means of Compliance (2009)
3. Stassen, H., et al.: Final Report of Control Engineering Group on Mental Load, in Mental Workload: Its Theory and Measurement, pp. 235–252. Plenum Press, New York (1979)
4. Eggemeier, F.T., et al.: Workload Assessment in Multi-Task Environments Multiple-Task Performance, pp. 207–216. Taylor and Francis, London (1991)
5. Kramer, A.F., Sirevaag, E.J., Braune, R.: A psychophysiological assessment of operator workload during simulated flight missions. Hum. Factors J. Hum. Factors Ergonomics Soc. **29**(2), 145–160 (1987)
6. Corwin, W.H., et al.: Assessment of Crew Workload Measurement Methods, Techniques and Procedures, vol. 1. Process, Methods and Results (1989). DTIC Document
7. Gawron, V.J.: Human Performance, Workload, and Situational Awareness Measures Handbook. CRC Press, Boca Raton (2008)
8. Di Nocera, F., Camilli, M., Terenzi, M.: A random glance at the flight deck: pilots' scanning strategies and the real-time assessment of mental workload. J. Cogn. Eng. Decis. Mak. **1**(3), 271–285 (2007)
9. Farmer, E., Brownson, A.: Review of workload measurement, analysis and interpretation methods. Eur. Organ. Safety Air Navig. **33** (2003)
10. Cain, B.: A Review of the Mental Workload Literature (2007). DTIC Document
11. Wickens, C.D., Huey, B.M.: Workload Transition: Implications for Individual and Team Performance. National Academies Press, Washington DC (1993)
12. Annett, J.: Subjective rating scales: science or art? Ergonomics **45**(14), 966–987 (2002)
13. Ramaekers, D., et al.: Heart rate variability and heart rate in healthy volunteers. Eur. Heart J. **19**, 1334–1341 (1998)
14. FAA, A.C., Airplane Simulator Qualification (AC 120–40B), F.A. Adminstration, (ed.), Washington, DC (1991)
15. Wickens, C.D.: Multiple resources and mental workload. Hum. Factors J. Hum. Factors Ergonomics Soc. **50**(3), 449–455 (2008)
16. Wickens, C.D., McCarley, J.S.: Applied Attention Theory. Taylor & Francis, Boca Raton (2007)
17. Buhusi, C.V., Meck, W.H.: Relative time sharing: new findings and an extension of the resource allocation model of temporal processing. Philos. Trans. R. Soc. B Biol. Sci. **364** (1525), 1875–1885 (2009)
18. Zheng, Y., et al.: Developing a measurement for task complexity in flight. Aerosp. Med. Hum. Perform. **86**(8), 698–704 (2015)
19. Kasarskis, P., et al.: Comparison of expert and novice scan behaviors during VFR flight. In: Proceedings of the 11th International Symposium on Aviation Psychology (2001)
20. Mumaw, R., Sarter, N., Wickens, C.: Analysis of pilot's monitoring and perfomance on an automated flight deck. In: International Symposium on Aviation Psychology (2001)