SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2024
ORGANIZATION AND MANAGEMENT SERIES NO. 200

THE USEFULNESS OF SUBJECTIVE TASK LOAD ASSESSMENT METHODS FOR PREDICTING PILOT TASK LOAD IN GENERAL AVIATION ORGANIZATIONS

Marcin BERLIK

Poznań University of Technology, Faculty of Engineering Management, Institute of Safety and Quality Engineering; marcin.berlik@put.poznan.pl, ORCID: 0000-0003-0365-7107

Purpose: The research aimed to verify the usefulness of selected subjective methods of task load assessment for evaluating task load in general aviation organizations and forecasting pre-flight load. The study also highlighted the importance of the pilot's task load in the context of flight safety.
Design/methodology/approach: The study included a comprehensive literature review to present the specifics of general aviation organization and the role of the human factor in aviation accidents. The research discussed the division of task load methods and characterized four selected subjective task load methods (NASA-TLX, SWAT, ISA, Bedford Scale). Expert assessments were conducted to evaluate these methods' effectiveness from the perspective of post-flight load assessment and pre-flight load prediction, considering the feasibility of implementation in general aviation organizations.
Findings: The findings emphasized the need to develop a dedicated method for predicting task load before a flight. The study concluded that the methods used should be simple and require minimal financial outlay, considering the specificity of general aviation organizations.
Research limitations/implications: Future research should focus on developing a dedicated pre-flight task load prediction method tailored to the unique requirements of general aviation. Identified limitations include the need for further validation of the subjective methods in different organizational contexts.
Practical implications: The research suggests that implementing simple and cost-effective task load assessment methods can enhance flight safety in general aviation. The development of a specialized pre-flight load prediction tool could significantly improve operational planning and pilot performance.
Social implications: By improving task load assessment and prediction in general aviation, the research can contribute to enhanced flight safety, thereby potentially reducing accident rates and improving public trust in aviation safety measures.
Originality/value: This paper contributes to the field by addressing the gap in pre-flight task load prediction in general aviation. It provides valuable insights for aviation safety researchers, practitioners, and policymakers interested in optimizing pilot performance and ensuring flight safety.
Keywords: safety management, operational risk management, flight safety, task load, human factor.
Category of the paper: Research paper.

1. Introduction

The aim of the research was to verify the usefulness of subjective methods of assessing task load from the point of view of small general aviation organizations.

According to the intention, the selected methods had to be used to assess the pilot's task load after the flight, but also to predict the task load before the flight, when it is possible. Task load prediction could constitute the basis for the assessment of operational risk related to the execution of a flight, which is practically not carried out at all in general aviation organizations.

The starting point for these considerations is the fact that approximately 80% of adverse events in aviation (accidents, incidents) are caused by human-related factors.

At the same time, from the point of view of safety culture, risk management and safety management, the greatest problems occur in general aviation.

2. Review of the literature

2.1. Specificity of the organization of general aviation

General aviation includes all aviation, excluding state aviation (mainly military aviation, police and border guard aviation) and commercial aviation. The majority of general aviation aircraft are aircraft owned by flight training organizations and private owners.

Especially in training flights, the task load has a significant impact on the operator's error (Hsu, Shu, Liu, Wang, 2022; Kang-Seok, Eun-Suk, Young, 2014; Romero, Robertson, Goetz, 2020; Szopa, 2015).

General aviation is the aviation sector with the highest number of fatal accidents. Despite many activities at international, European and national levels, and the increased efforts made by general aviation organizations, these trends are continued. General aviation is a very specific sector; it is characterized by a relatively low safety culture, the need to manage safety using a small budget, high pilot turnover, and often an outdated or heterogeneous fleet. A large proportion of flights undertaken in this sector are recreational or training flights, which carry a higher risk - because they are often made by pilots with relatively little experience, as an example. Also, training flights, due to their specificity, which includes the need for intensive flight practice in dangerous situations by student pilots on their own, as an example, involve increased risk (Sun, Yang, Zhang, Zhao, 2023).

Among the causes of high-risk events in the area of human factors, the most common causes include task performance, situational awareness and sensory events (figure 1).

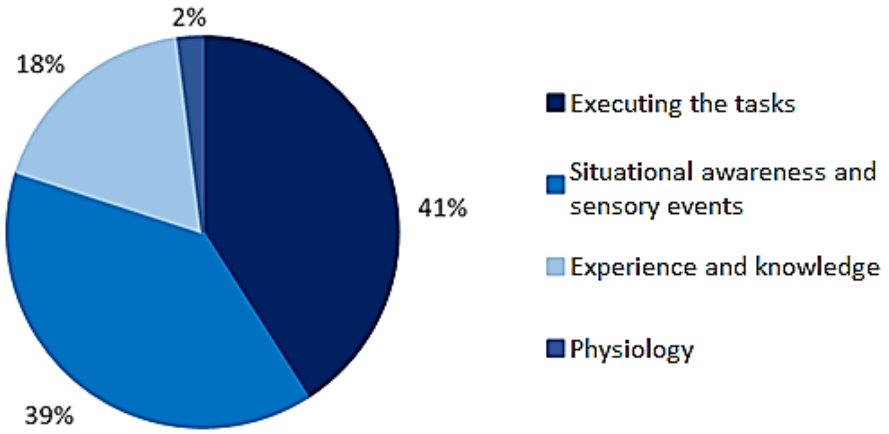

Figure 1. Division of high-risk events in the area of human factors, according to causal codes for accidents and serious incidents. High-level event codes related to human factors and human performance, applicable to accidents and serious incidents.

Source: own study based on research results EASA, 2022.

These data indicate a significant impact of task load management on flight safety.

2.2. Division of task load methods

There are many methods for assessing task load. They are most often used to determine whether an operator is overloaded or underloaded; whether there is a need to reduce the load; whether there are still resources available to be used for additional tasks (Masi, Amprimo, Ferraris, Priano, 2023). Assessment results can contribute to actions aimed at the reduction of the load, by providing support to the person through technical or organizational means (Davis et al., 2014; United States Department of Transportation, 2019).

Despite the availability of many methods, assessing task load is a difficult task. This is due to the limitations of each of the methods developed so far and the presence of many variables that affect human functioning (Casper, Kantowitz, 2009). Selected methods, that are considered particularly useful from the point of view of their application in aviation, are characterized below. They are divided into subjective, semi-objective and objective (Alaimo, Esposito, Orlando, Simoncini, 2020; Berlik, Ewertowski, Sławińska, 2019).

Subjective methods include methods that are based on a person's own assessment. The pilot assesses the load while performing the task by providing answers using a questionnaire. The advantage of using these methods is obtaining a person's subjective opinion about his or her experiences. However, the possibility of manipulating the research results by filling out the questionnaire unreliably is its disadvantage.

Semi-objective methods, also based on human assessment, include all methods in which the assessment is made by the person conducting the observation. In the case of pilot's load evaluation, the assessment may be made by the training or controlling the instructor. Another example of an assessor is a so-called "safety pilot", i.e. a pilot who is not in command of the aircraft and who i.e. accompanies a less experienced pilot to enable the response to emergency situations. Similarly to subjective methods, semi-objective methods also involve a limited

perception of the situation by the assessor and the risk of manipulation which is made by him, but such an assessment is largely objectified because it is performed by a third party, who usually has greater knowledge and experience.

The last group consists of objective methods, which include all methods based on the recording of human physiological parameters. Their advantage is a very limited scope of interference by the examined person on the obtained measurement results. Unfortunately, these methods are expensive, and the interpretation of the results is simultaneously very demanding, and their implementation in general aviation organizations is very difficult or often impossible. For this reason, this study focuses on subjective and semi-objective methods.

2.3. A review of subjective and semi-objective task load methods

Subjective Workload Assessment Technique (SWAT) (SWAT Eurocontrol)

SWAT is a multidimensional workload assessment method for pilots that takes into account the following categories (Zhang, Zheng, Duan, Meng, Zhang, 2015):

- time deficit,
- mental load,
- stress level.

The SWAT assessment consists of two stages. The first one, and the most time-consuming one, involves developing a scale. The subject arranges 27 cards in the order corresponding to how the workload is perceived by the subject - from the smallest to the largest. Based on the respondent's answers, an analysis is carried out to develop an individual scale. In the second stage, the subject assesses the workload for a given task. This is done based on the assessment of specific activities according to the indicated categories (time deficit, mental load, stress level). In each of these areas, the respondent can rate the level of load on a scale of one to three points. Then the results are converted in accordance with the scale developed in the first stage. The end result is an assessment of the load level in the form of a numerical result ranging from zero to one hundred points.

National Aeronautics and Space Administration Task Load Index (NASA-TLX)

NASA-TLX is a questionnaire method developed by the Human Performance Group at NASA's Ames Research Center (Zhang et al., 2015). During the test, using a prepared questionnaire, the task load is assessed after the flight in six indicated categories:

- mental load,
- physical load
- time pressure,
- efficiency,
- effort,
- frustration.

During the assessment, the tested pilot marks the appropriate part of the twenty-point scale on the questionnaire, corresponding to his load range. Expressing the level of load by marking on a scale, rather than providing a numerical value, is intended to provide a more reliable evaluation. Then, in a choice test, where two categories are contrasted, the subject selects the one that generates the feeling of the greatest load (Berlik, Ewertowski, Sławińska, 2019; NASA-TLX, 2009).

The advantage of using the NASA TLX method is that it is quite easy to conduct the test. Completing the questionnaire (Fig. 2) takes a few minutes, and analysing the answers allows you to obtain sufficiently accurate results. The great advantage of the NASA TLX method is the identification of the categories in which the greatest load occurs. There are many applications available (Fig. 3) for iOS and Android systems, as well as spreadsheet templates that enable quick preparation of research results (Gawron, 2008).

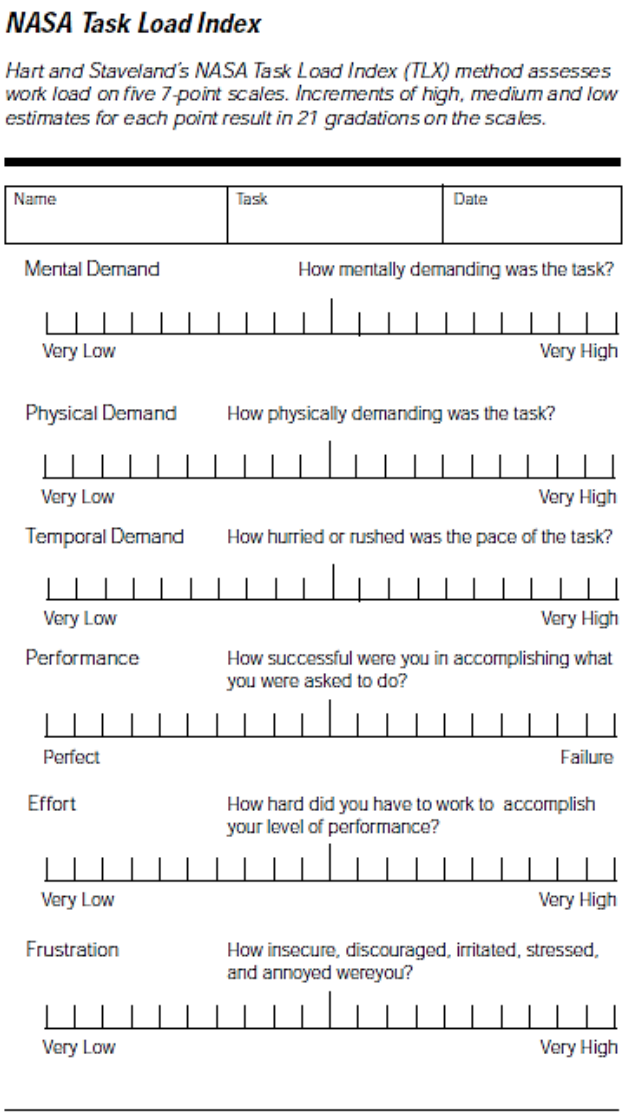

Figure 2. NASA TLX questionnaire.

Source: NASA TLX, 2009.

Figure 3. Sample result of an assessment conducted using the NASA-TLX questionnaire in the application.

Source: NASA TLX App Store.

Bedford Workload Scale

In the Bedford Workload Scale, the subject assesses his or her workload on a ten-point scale (Figure 4) using auxiliary questions. Thanks to the auxiliary questions regarding whether the task was possible at all, whether the level of load was tolerable and whether the level of load was satisfactory without the use of additional tools to reduce it, detailed questions are separated to which the appropriate level of load is assigned (Jennings, Craig, Carignan, Ellis, Qinetiq, 2005). The big advantage of this method is the significantly shorter evaluation time than other methods and the fact that the results do not require detailed analysis. Unfortunately, the assessment carried out using this method does not allow to identify the categories in which the load was greatest. However, it can be a method of quickly assessing whether a person is overloaded, underloaded, or whether the load is adequate. Thanks to this, it is possible to suspend the implementation of the task and perform a detailed diagnosis (Berlik, Ewertowski, Sławińska, 2019; Sgobba et al., 2017).

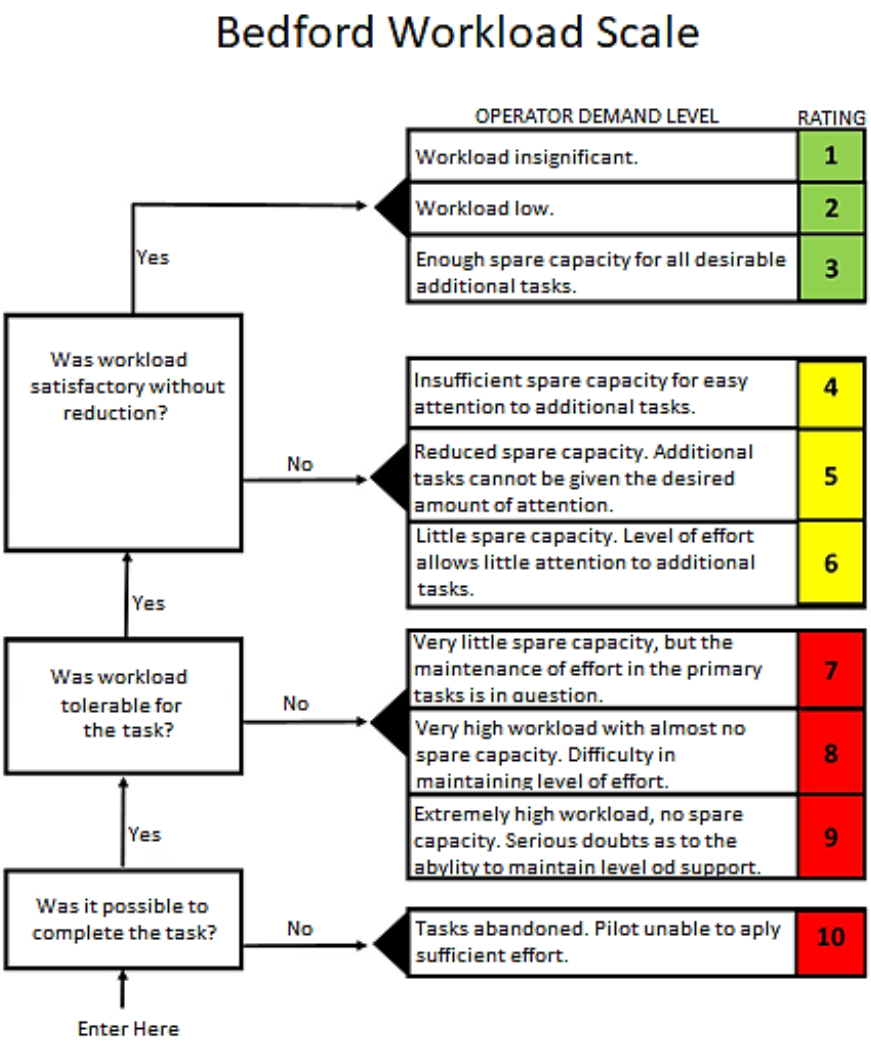

Figure 4. Bedford Workload Scale.

Source: own study based on research results Mari, et al., 2014.

Instantaneous Self Assessment (ISA)

ISA is a technique for immediately assessing the extent to which an operator's resources are committed to performing tasks (Hamann, Carstengerdes, 2020). The assessment is made by selecting the appropriate key on the keyboard located at the workstation. The subject assesses the degree of load by selecting a key in the appropriate colour, additionally marked with a description. The operator is informed about the need to perform an assessment by the illumination of a diode, that is located at his eye level. What is characteristic of this method is that the assessment is made while performing the task (Radüntz, Fürstenau, Tews, Rabe, Meffert, 2019). A description of the individual levels of load assessment is provided in Table 1.

Table 1.
Instantaneous Self Assessment (ISA) stress levels

Rating	Load	Description
1	Unused	Nothing to do, boredom
2	Relaxing	More than enough time to complete all tasks. Task-related activity for less than 50% of the total expected time
3	Comfortable	All tasks possible to complete. Brisk and stimulating pace of load development. Work is still possible at this level
4	High	Insignificant difficulties in performing tasks. Work at this level cannot be performed for too long
5	Excessive	Delays in completing tasks. Loss of full picture of activities performed

Source: own study based on Rusnock, Geiger, 2017.

The ISA method is most often used in examining the work of air traffic controllers. It would be difficult to use it in the study of a pilot's work during a flight, especially in the key phases that are most interesting from the researcher's point of view and generate the greatest load - such as take-off or landing. However, there are known cases of this method being used in a flight. An example would be research conducted during experimental flights conducted by Airbus. During the tests, a keyboard was mounted to the aircraft's instrument panel. The diode was activated manually by an observer on board when the nature of the tasks performed changed significantly or at a standard three-minute interval. In addition, the observer assessed the load level of both pilots performing the flight on a seven-point scale and recorded the errors made on a three-point scale, identifying them as: minor, significant and having an impact on safety.

Due to the need to provide answers while performing the task, this method is useful from the point of view of examining the pilot's load in the simulator. Unfortunately, the impact of the conducted research during the flight may be too great on its safety (ISA, 2023).

Research is also carried out to develop other methods of assessing task load (especially objective methods). Unfortunately, due to their lack of availability or lack of confirmed effectiveness, they were not included in this study. It is possible that over time, thanks to intensive research and development work, in the near future they will become available and useful from the point of view of measuring task load and managing flight safety (Berlik, 2023).

3. Research methodology

As part of the selection of tools for assessing the pilot's task load, the selected tools were tested during flights. The verified tools were selected on the basis of a focused interview and the results of an assessment conducted by experts in 2020. The following describes the interview and expert assessment, as well as the experience gained during the test flights.

When selecting the experts, it was assumed that they should have extensive aviation experience, especially experience in training, observing and assessing pilots' behaviour in various workload situations. Based on these assumptions, the following requirements were adopted:

- possession of a pilot license,
- possession of a valid general training instructor certificate (formerly 1st class) authorizing the student pilot to perform the first solo flight and the first solo flight/route flight in his/her life,
- having at least 400 hours of general flight time,
- having at least 100 hours of flight time as an instructor.

The interview was helped by the fact that the author, being a pilot and instructor himself for many years, met the above requirements and well understood the issues raised.

The experts' task was to assess the usefulness of the proposed tools from the point of view of practical application in a small general aviation organization.

Instructors were to take into account aspects such as:

- the possibility of conducting a comprehensive load assessment, also including technical, environmental, management and procedural factors affecting the pilot's load;
- feasibility of using the solution in the organization. The following was considered: tendency of pilots and instructors to perform additional tasks and the possibility of analysing data obtained with means of tools and by people operating in the organization;
- functionality and possible difficulties resulting from the use of tools, especially during flight;
- the impact of the applied tools on the safety of flights.

Instructors rated the tools on a scale of 1 to 5 points, where 1 point meant the least usefulness and 5 points the most usefulness.

An average score of 2.5 points or more was considered satisfactory.

4. Research results

The results of the evaluation of individual tools are presented in the table below (Table 2). In the instructors' assessment, the NASA-TLX and Bedford Workload Scale methods achieved a score of 2.5 points and higher. The most useful tool was the NASA-TLX questionnaire (4.5 points), and the second tool - the Bedford task load questionnaire - was rated at 3.75 points.

Table 2.
Results of the method usefulness assessment for measuring the pilot's task load

Results of the method usefulness assessment for measuring the pilot's task load					
	instructor 1	instructor 2	instructor 3	instructor 4	average rating
Subjective Workload Assessment Technique (SWAT)	2	1	1	2	**1,5**
National Aeronautics and Space Administration Task Load Index (NASA-TLX)	5	3	5	5	**4,5**
Bedford Workload Scale	5	1	4	5	**3,75**
Instantaneous Self Assessment (ISA)	1	1	2	1	**1,25**

Source: own study based on (Berlik, 2023).

In their comments, instructors mainly indicated practical aspects. They found that the NASA-TLX and Bedford Workload Scale methods are the most useful from the point of view of a small general aviation organization, because their application is the simplest and does not require the installation of additional devices and long-term familiarization of the pilot with the assessment principles. At the same time, it was demonstrated that tools based on questionnaires are useful in methods that build awareness of the risks and difficulties that may be encountered during the flight, when the pilot, who is familiarised with the assessment method before the flight, will be able to interpret the level of his skills and preparation for the flight and compare this knowledge with his thoughts on the difficulties related to the performance of the task. This is especially important for less experienced pilots and may also contribute to better preparation for the flight. At the same time, concerns were expressed that an inexperienced pilot may not be aware of potential difficulties during the flight or may even not know what factors may affect the load. It was suggested that the use of auxiliary questions, to which the pilot should answer before the flight, could help the pilot take a comprehensive insight into various aspects of the planned task and compare own capabilities with the conditions of the executed task (Berlik, 2023).

NASA-TLX

It was demonstrated in the experts' assessment, that it may be quite a problem to explain to respondents the meaning of the names of individual categories.

Also during the tests, it turned out that the simplicity of the method was its drawback. It was found that a major problem may be not only explaining to respondents the meaning of the names of individual categories, but also the load factors occurring in them.

Despite the above problems, completing the post-flight survey was not difficult. Unfortunately, pre-flight load prediction using this method was difficult. This was due to the fact that particularly the inexperienced pilots had trouble understanding the factors that could generate load in given categories. They did not take into account many aspects and needed advice on what load-related factors may occur during the flight.

The above problems contributed to the development of own questionnaire, based on the areas identified by NASA-TLX. It consisted of specific questions about the various factors that generate load during and before flight on a given day, organized into 6 NASA-TLX load-related groups.

Bedford Workload Scale

The tests demonstrated the possibility of carrying out the method very quickly, as well as its convenience in assessing the load during the flight or its selected phase (e.g. landing). During the tests, the pilot was also asked to indicate the phase of flight in which he felt the greatest load and to rate it on the Bedford scale. In the case of the research conducted, the method was considered unusable due to its poor usefulness in terms of load forecasting.

5. Conclusions

The analyses carried out and the obtained research results indicate the greatest usefulness of the NASA-TLX method for the purposes of assessing the pilot's task load in general aviation organizations. However, the use of this method may be difficult due to the unclear definition of the six areas of task load identified in the questionnaire.

A much easier and less time-consuming method is the Bedford Workload Scale. The one-dimensional result obtained in this method can be an advantage because it clearly helps to determine to what extent the task performed was adequate to the level of knowledge and skills of the pilot. An important advantage may also be the ease of answering by the respondent.

Unfortunately, in the author's opinion, the usefulness of all analysed methods for pre-flight load prediction is low. This problem particularly concerns pilots with little experience, who have little awareness of the factors affecting the task load during a flight and for whom making a prediction may be extremely difficult, and it cannot be ensured that such a prediction will be made in a comprehensive manner.

According to the author, it is advised to develop a special questionnaire dedicated to predicting task load before a flight, which would take into account individual task load factors. Completing such a questionnaire, which included detailed questions about the factors affecting the task load associated with the flight, could facilitate accurate prediction of the load, but also increase the pilot's awareness of the risks associated with the flight and be helpful in self-

monitoring the flight preparation. The questions included in the questionnaire should concern various issues related to the performance of the task by the pilot, such as: weather conditions, the type of space in which the flight will be performed and the procedures applicable there, knowledge of the aircraft and the difficulty of piloting it, the pilot's well-being, etc.

The development of such a questionnaire should be preceded by literature research, and its usefulness should be verified during the research. Due to the specificity of general aviation organizations, which usually employ only a few people, have a small budget and where the pilot is also the customer, it is particularly important to select the methods that are as least burdensome as possible. Some of the most important criteria when developing it should be: low time consumption related to both its completion and analysis of results, simplicity and low cost (Berlik, 2023).

6. Summary

This article discusses the importance of the human factor in managing flight safety, as well as the importance of the pilot's task load in this area. The basic limitations faced by general aviation organizations in managing flight safety and monitoring the pilot's task load were also discussed.

Based on the analysis of four selected subjective methods, which can also be used as semi-objective methods, their usefulness in assessing the task load after a flight was noticed. Unfortunately, they turned out to be of little use in terms of the pilot's pre-flight load prediction. This is especially a problem for young, less experienced pilots who are less aware of the factors that may occur during flights. It is advisable to develop questionnaire methods containing detailed questions regarding pilot load factors during the flight. They should take into account as comprehensively as possible the factors affecting the pilot's task load and refer not only to the person himself, but also to the fullest extent possible to the elements of technology or the environment influencing the performance of the aviation task. The use of questionnaires constructed in this way could enable the prediction of the pilot's task load during the flight, and thus the prediction of the operational risk associated with the flight. Such questionnaires could also contribute to the increased level of safety of air operations by increasing the pilot's awareness of the task being performed and enabling self-monitoring of his or her preparation for the flight.

References

1. Alaimo, A., Esposito, A., Orlando, C., Simoncini, A. (2020). Aircraft Pilots Workload Analysis: Heart Rate Variability Objective Measures and NASA-Task Load Index Subjective Evaluation. *Aerospace, 7*(9), 137.
2. Berlik, M. (2023). *Metoda zarządzania ryzykiem operacyjnym na podstawie obciążenia zadaniowego pilota w bezpieczeństwie lotów* (Doctoral dissertation). Poznań University of Technology.
3. Berlik, M., Ewertowski, T., Sławińska, M. (2019). Przegląd metod oceny obciążenia zadaniowego w aspekcie doskonalenia układów operator – obiekt techniczny na przykładzie pilota. *Journal of KONBiN, Vol. 49, sheet 3*, 97-114.
4. Casper, P., Kantowitz, B. (2009). *Human Workload in Aviation.*
5. Davis, B. et al. (2014). Effective Teamwork and Communication Mitigate Task Saturation in Simulated Critical Care Air Transport Team Missions. *Military Medicine, 179*, 19-23.
6. European Union Aviation Safety Agency - European Union Aviation Safety Agency - EASA (2022). *Annual Safety Review*.
7. Gawron, V.J. (2008). *Human Performance, Workload, and Situational Awareness Measures Handbook.* Boca Raton: CRC Press.
8. Hamann, A., Carstengerdes, N. (2020) *Fatigue Instantaneous Self-Assessment (F-ISA): Development of a Short Mental Fatigue Rating*. Retrieved from: https://elib.dlr.de/135577/, 26.08.2023.
9. Hsu, W.-K. K, Shu, M.-H., Liu, Y.-C., Wang, T.-C. (2022). Risk Management of Safety for Flight Training in Air Forces. *Aerospace, 9(10),* 558.
10. ISA Human Factors Methods (2023). Retrieved from: https://www.skybrary.aero/articles/instantaneous-self-assessment-isa, 3.07.2023.
11. Jennings, S., Craig, G., Carignan, S., Ellis, K., Qinetiq, D.T. (2005). Evaluating control activity as a measure of workload in flight test. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 49*(1), 64-67.
12. Kang-Seok, L., Eun-Suk, S., Young, S. (2014). Impact of human factors for student pilots in approved flight training organizations in Korea. *Aviation / Aeronautics / Aerospace International Research Conference, 45*, 236-249.
13. Maris, J.M., Haynes, M.A., Wing, D.J., Burke, K.A., Henderson, J., Woods, S.E. (2014). *Traffic Aware Planner (TAP) Flight Evaluation.* Retrieved from: https://ntrs.nasa.gov/api/citations/20140011404/downloads/20140011404.pdf, 30.08.2023.
14. Masi, G., Amprimo, G., Ferraris, C., Priano, L. (2023). Stress and Workload Assessment in Aviation—A Narrative Review. *Sensors, 23*(7), 3556.

15. *Nasa Task Load Index – NASA-TLX* (2009). Retrieved from: https://humansystems.arc.nasa.gov/groups/tlx/downloads/TLX_pappen_manual.pdf, 26.06.2023.
16. *NASA TLX App Store*. Retrieved from: https://itunes.apple.com/us/app/nasa-tlx/id1168110608?mt=8 (access: 26.06.2023)
17. Radüntz, T., Fürstenau, N., Tews, A., Rabe, L., Meffert, B. (2019). The Effect of an Exceptional Event on the Subjectively Experienced Workload of Air Traffic Controllers. In: L. Longo, M. Leva (eds.), *Human Mental Workload: Models and Applications.*
18. Romero, M.J., Robertson, M.F., Goetz, S.C. (2020). Fatigue in Collegiate Flight Training. *The Collegiate Aviation Review International, vol. 38, No. 1*.
19. Rusnock, Ch.F., Geiger, Ch.D. (2017). Designing adaptive systems: selecting an invoking threshold to improve human performance. *International Journal of Human Factors and Ergonomics, 4*, 292-315.
20. Sgobba, T. et al. (2017). *Space Safety and Human Performance.* Cambridge: Butterworth-Heinemann
21. Sun, H., Yang, F., Zhang, P., Zhao, Y. (2023). Flight Training Risk Identification and Assessment Based on the HHM-RFRM Model. *Sustainability*, *15*, 1693.
22. Szopa, T. (2015). *Niezawodność i bezpieczeństwo*. Warsaw: Publishing House of the Warsaw University of Technology.
23. United States Department of Transportation. (2019). Workload Assessment, Retrieved from: https://railroads.dot.gov/human-factors/elearning-attention/workload-assessment. (access: 5.07.2023)
24. Zhang, Y., Zheng, H., Duan, Y., Meng, L., Zhang, L. (2015). *An integrated approach to subjective measuring commercial aviation pilot workload.* IEEE 10th Conference on Industrial Electronics and Applications (ICIEA). Auckland, New Zealand, 1093-1098.