# A Multidimensional Method for Pilot Workload Assessment and Diagnosis

Zhen Wang; Yanyu Lu; Shan Fu

**INTRODUCTION:** Pilot workload assessment plays an important role on flight safety evaluation, interface design, and airworthiness certification. The design of an effective and reliable workload assessment method is a difficult problem in the human factors field.

**METHOD:** This study proposed to assess pilot workload from four dimensions: cognitive activity, control activity, stress, and flight performance. A set of physiological, behavioral, and flight parameters were recorded and combined hierarchically to achieve overall workload assessment. A simulated flight experiment consisting of three flight phases (standard instrument departure, autopilot cruise, and nonprecision approach) was conducted to test the effectiveness of the proposed workload assessment method.

**RESULT:** Experimental results determined the changes of each objective measure. The overall workload index could significantly distinguish the difference in workload caused by changing task difficulty and the result was consistent with the NASA-TLX. The four workload dimensions provided detailed differences about workload: during nonprecision approach there were more control activities and stress than in other flight phases; during autopilot cruise there were the least control activities and the highest flight performance. The correlation between workload dimensions provided extra diagnostic information: the cognitive and control activities in the approach phase were more stressful than in the takeoff phase; the correlation between control activity and performance was higher in the approach phase than in the takeoff phase.

**CONCLUSION:** This study proposed an effective pilot workload assessment method which could also provide detailed and diagnostic information. It could be used as an auxiliary tool for the development and evaluation of pilot-cockpit interaction.

**KEYWORDS:** pilot workload, multidimensional assessment, simulated flight, diagnostic, objective measures.

**Wang Z, Lu Y, Fu S. *A multidimensional method for pilot workload assessment and diagnosis*. Aerosp Med Hum Perform. 2020; 91(12):932–939.**

Up to 70% of civil aircraft accidents are related to human errors.[11,22] One of the major causes of human error is excessive workload imposed on pilots.[17] Workload can be affected by various factors, e.g., task demand, cockpit interface designs, expertise of the pilots, etc.[7] Comprehensive and reliable assessment of pilot workload has great importance to the evaluation of flight safety, as well as to the development and certification of aircraft cockpits.

Conventionally, workload assessment techniques can be classified into three categories.[23] 1) Performance measures. According to the resource-performance function, a more demanding task requires more human resource and also results in less remaining resource.[28] When task performance (primary task performance or secondary task performance) degrades, the changes of workload can be reflected. 2) Physiological measures. They are based on the fact that the autonomic nervous system regulates the activity of organs when circumstances change. Various physiological indices have shown sensitivity to workload.[3,25,30] Also, physiological parameters are generic measurements that could be applied in different fields, and they could be recorded continuously and provide detailed information for the entire task. 3) Subjective rating scales. These use the operator's own experience to reflect workload. Several subjective techniques have been validated in various areas.[1] Subjective

From the Department of Automation, Shanghai Jiao Tong University, Shanghai, China.

This manuscript was received for review in January 2020. It was accepted for publication in September 2020.

Address correspondence to: Shan Fu, Department of Automation, Shanghai Jiao Tong University, No. 800 Dongchuan Rd., Minhang District, Shanghai 200240, China; sfu@sjtu.edu.cn.

Reprint & Copyright © by the Aerospace Medical Association, Alexandria, VA.

DOI: https://doi.org/10.3357/AMHP.5592.2020

http://prime-pdf-watermark.prime-prod.pubfactory.com/ | 2025-02-05

rating scales are simple and convenient to perform. They are the most widely used workload measuring techniques.

However, there are some inherent shortcomings in the existing workload assessment techniques. For example, the primary task performance measures overlook the compensatory effort invested by the operator.[4] It is not dependable when used alone. The secondary task has to compete against the primary task for human resources from the same resource pool.[23] For a complex primary task which requires multiple resources, it is extremely difficult to set up a satisfactory secondary task. Moreover, a secondary task would interfere with primary task performance and become the source of workload itself. For physiological measures, the sensitivity of physiological indices is often task dependent.[24] The explanations of physiological parameters are often ambiguous.[5] A single physiological index would be insufficient to provide a reliable workload assessment.[15] For subjective measures, they can hardly be performed continuously.[23] When performed after the task, the participant might forget their feelings in some important moment. When performed intermittently during the task, they may affect the participant's task performance. Moreover, participants would find the questionnaire items confusing if they do not have human factors research experience.

A way to improve general sensitivity and reliability is to integrate a battery of indices. Researchers have proposed a number of methods to combine different workload indices, such as subjective weighting,[19] multiple regression analysis, and machine learning method, e.g., Bayes network,[2] ANN,[32,33] SVM,[6] etc. The integrated methods are thought to be more reliable than individual measurement because they consider the concordant and redundant information among the indices, and the models were often determined with large data sets. However, there are still some problems in these integrated methods. For the subjective weighting method, their validity largely depends on the number and experience of the experts. For the multiple regression and supervised machine learning method, the weightings or the structure of the model depend on a large amount of training data. However, in some areas, such as aviation, it is extremely difficult to acquire a training data set with satisfactory sample size. Furthermore, combining multiple measurements directly into a single index would lose a lot of useful information. It is hard to extract diagnostic cues from a unidimensional result.[8]

In this study, we stated our understanding of workload, proposed a multidimensional workload assessment and diagnosis method, and tested the proposed method in a simulated flight experiment. From our perspective, pilot workload is interpreted as the pilot's interactive activities with cockpit components in achieving a particular level of performance and under a certain level of stress.

Specifically, a pilot's interactive activities involve two major aspects. In one aspect, pilots have a large amount of cognitive activities to discern cues from the cockpit, evaluate the situation, and make decisions. In another aspect, pilots have a series of control activities to provide inputs to the cockpit components to change the status of the aircraft. More task demands often requires more cognitive activities and control activities.

Increasing levels of task demand would not only affect a pilot's interactive activities with the cockpit, it would also put more stress on the pilot. Sometimes, even with the same operational procedure and task demand, the stress imposed on pilots could be different. Yao et al.[34] has indicated that during the takeoff phase, less experienced pilots were under higher stress than more experienced pilots. Stress is one of the most important aspects of workload.

When an aircraft is deviating from the route, but the pilot still looks quite relaxed and does not take necessary actions to correct the deviation, it could be deduced that the pilot is too careless to maintain situation awareness and his workload is not good for flight safety. Thus, just a statement of the pilot's status is meaningless unless flight performance is also taken into account.[13]

In this study, workload was considered a multidimensional construct. We proposed to assess pilot workload from four dimensions: cognitive activity (CGN), control activity (CTRL), stress (STRS), and flight performance (PFM). To acquire satisfactory sensitivity and reliability, each workload dimension is described by integrating multiple objective parameters.

For the cognitive activity dimension, fixation duration, blink interval, and saccade rate are selected and integrated. Yu et al.[35] indicated that fixation duration can reflect a pilot's information processing pattern and situation awareness performance. Veltman et al.[24] pointed out that blink interval increased as more visual information had to be processed. Nakayama et al.[16] indicated that saccade rate decreased when there was higher demand for information processing.

For the control activity dimension, the deflection speed (angular changes between adjacent sample points) of the control column, control wheel, pedals, and throttle lever are integrated to reflect pilot's controls to the aircraft.[29] For the pilot stress dimension, heart rate,[9,12,34] respiration rate,[24,34] respiration amplitude,[25] and pupil diameter[18,20] were selected and integrated. These physiological indices have shown certain sensitivity to an operator's stress in previous studies. For the flight performance dimension, lateral deviation, altitude deviation, and airspeed deviation between actual flight and the flight plan were selected and integrated.

The integration of multiple parameters is carried out with the following steps:

- Synchronize all the measured date according to their sample timestamps.
- Normalize the indices by calculating their z-scores.
- For each workload dimension, analyze the relevant objective parameters with principal component analysis. Principal component analysis considers the variation of each parameter and the relationships among the parameters. It can reveal the principal information included in a multivariate structure. Then multiple parameters are integrated with Eq. 1:

$$w = \sum_{i=1}^{n} a_i \cdot x_i \tag{1}$$

http://prime-pdf-watermark.prime-prod.pubfactory.com/ | 2025-02-05

where $w$ represents the value of a workload dimension. If the workload dimension is described by $n$ objective parameters, then there would be $n$ principal components. $a_i$ is the proportion of the variance explained by the $i^{\text{th}}$ component during the task. $x_i$ is the score of the $i^{\text{th}}$ component.

- An overall workload assessment index is also synthesized from the four workload dimensions via principal component analysis and Eq. 1. This overall assessment index can help the evaluators to make general workload comparisons.

It is important for a workload assessment method to have diagnostic capability, because the ultimate aim of workload assessment is to find out the cause of inappropriate workload and provide guidance to system design, operational procedure optimization, and personnel training.[28] In this study, we not only studied workload assessment from multiple dimensions, we also supposed that these dimensions are not independent, and the interactions among them could reveal additional diagnostic information. For example, by comparing the CGN-STRS correlation with the CTRL-STRS correlation, we might infer which kind of activity is more likely to put stress on a pilot. From the CGN-PFM correlation, we might infer whether changes in flight performance catch a pilot's attention. From the CTRL-PFM correlation, we might infer whether a pilot's controls are effective. Pearson product-moment correlation coefficients between the workload dimensions are calculated. The absolute value of the correlation coefficient is used to describe the relationship between relevant workload dimensions.

It has to be noted that in the model of human information processing stages,[28] information perception and control execution are not directly linked; neither are attentional resources occupancy (stress) and system feedback (performance). Thus, the diagnostic meanings of CGN-CTRL correlation and STRS-PFM correlation are not so straightforward. These correlations are not considered in this study.

So far, the structure of the proposed pilot workload assessment method can be illustrated with a multidimensional pattern as in **Fig. 1**. In this pattern, the pilot-cockpit interactions (cognitive activity and control activity) lie in the vertical direction; pilot's stress and flight performance lie in the horizontal direction. The diagonal directions represent the correlations between adjacent workload dimensions. This octagon pattern provides an intuitive description of pilot workload.

## METHODS

### Subjects

To test the proposed workload assessment method, 10 male graduate students from Shanghai Jiao Tong University (mean age = 24.6, SD = 2.2) took part in a simulated flight experiment. None of the subjects had any flight experience before. None of them suffered any illness or took medications at the time. All the subjects were informed of the purpose and procedure of the experiment and signed informed consent forms before participation. The research was approved by the Institutional Review Board of Shanghai Jiao Tong University (Approval number: 2,017,033).

### Apparatus

The experiment was performed in a Boeing 777-200ER flight simulator which was built up based on FlightGear. The simulator can record the flight data (e.g., longitude, latitude, attitude, airspeed, etc.) and the control data (e.g., position of the control column, control wheel, pedals, throttle lever, etc.) in real time with a sample rate of 30 Hz. It has to be noted that the simulator did not have a nose wheel tiller; therefore, the pedals were used to control both the rudder and front wheel.

Two sets of physiological status monitors were deployed to record participants' physiological parameters. The SmartEye desk mounted eye tracker (Smart Eye AB, Gothenburg, Sweden) was used to capture each participant's pupil and gaze with a sample rate of 60 Hz. Blink interval, fixation duration, saccade rate, and pupil diameter were then extracted from the raw data.

The Bioharness system (Bioharness, Zephyr Technology Corp., Annapolis, MD) consists of a sensor embedded chest strap and a portable data acquisition module. It measures heart rate from the ECG signal at a sample rate of 250 Hz. By detecting the size differential of the thorax with the sample rate of 18 Hz, respiration rate and amplitude (voltage on the pressure sensitive sensors) can also be detected.

### Procedure

The simulated flight task consisted of a complete flight from takeoff to landing. The aircraft would take off from the San Jose International Airport (ICAO code: KSJC), runway 30R. After passing five waypoints, it would land at San Francisco International Airport (ICAO code: KSFO), runway 28R. The top view of the flight profile is illustrated in **Fig. 2**.

The flight task can be divided into three phases: takeoff, cruise, and approach landing. The takeoff phase started when the aircraft began to taxi along the runway and ended at waypoint SUNNE. In this phase, participants were required to perform the standard instrument departure. The cruise phase started from waypoint SUNNE and ended at waypoint CEPIN. In this phase, participants were asked to engage autopilot (LNAV + VNAV mode) and auto-throttle so that the aircraft was completely controlled by automatic systems. Meanwhile, the function of the participants was to monitor the automatic systems and maintain situation awareness. The approach landing phase started from waypoint CEPIN and ended when the aircraft had landed and stopped on the runway. In this phase, autopilot and auto-throttle were disengaged by the experimenter, and participants had to perform a nonprecision approach with manual control. Therefore, the task difficulty level for different flight phases would be: approach landing $>$ takeoff $>$ cruise.

There was a training session before the experiment for participants to familiarize themselves with the simulator and procedure. During this session each participant accumulated 6 h of flight experience.

http://prime-pdf-watermark.prime-prod.pubfactory.com/ | 2025-02-05

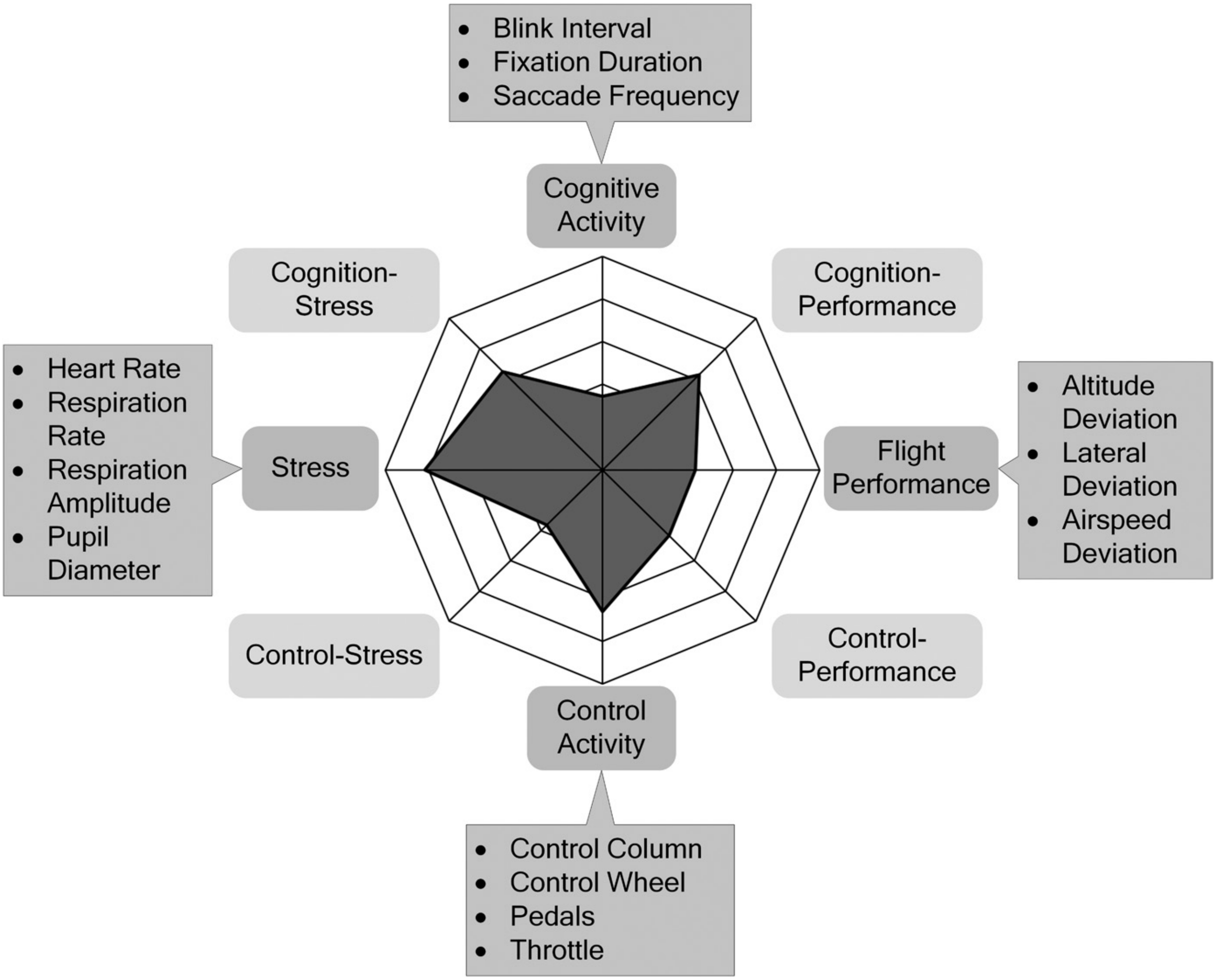

**Fig. 1.** The framework of the proposed workload assessment and diagnosis method.

During the formal experiment, after device calibration and testing, there was a 10-min rest period for participants to relax. Then the experimenter started the simulation and all the objective parameters were recorded concurrently.

When each flight phase ended, the experimenter froze the simulator and paused recording. Participants had 5 min to complete the NASA-TLX subjective rating scales followed by a 5-min rest. After that, the experimenter unfroze the simulator and resumed recording. Participants proceeded to perform the next task in the next flight phase.

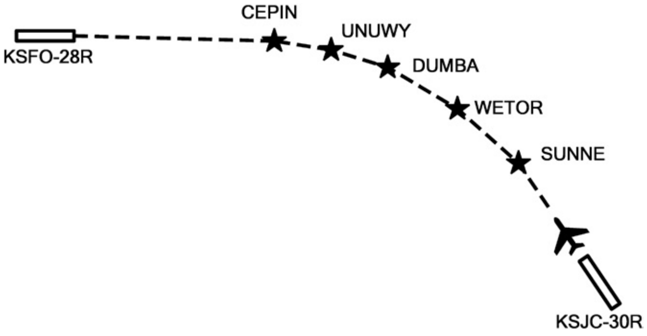

**Fig. 2.** Top view of the flight profile.

### Statistical Analysis

Statistical analyses were performed to test the means of the indices among different flight phases. Specifically, one-way ANOVA (analysis of variance) was applied to test the means of NASA-TLX scores among different flight phases.

For the objective measurements and the synthetic parameters (workload dimensions and overall workload assessment index), MANOVA (multivariate analysis of variance) was applied to test the multivariate population means among different flight phases. It is a good option to use MANOVA when there are more than two dependent variables because it considers the intercorrelations of the dependent variables, and it is robust to minor violations of the normality assumption.[10]

http://prime-pdf-watermark.prime-prod.pubfactory.com/ | 2025-02-05

Additionally, the means of each parameter among different flight phases was examined by one-way ANOVA. For correlation analysis, the absolute value of the Pearson product moment correlation coefficients (and the corresponding $P$-value) between CGN and PFM, CTRL and PFM, CTRL and STRS, and CGN and STRS were calculated in each flight phase, respectively.

## RESULTS

The NASA-TLX scores showed significant difference among the three flight phases [$F(2, 27) = 6.65$, $P = 0.004$]. A Tukey post hoc test revealed that the experienced workload was significantly higher in the approach landing phase (46.57 ± 14.56) than in the takeoff phase (26.90 ± 13.48, $P = 0.005$) and cruise phase (31.12 ± 9.51, $P = 0.029$). The difference between the takeoff and cruise phases was not significant ($P = 0.740$).

For objective measurements and their synthetic parameters, the MANOVA result showed that there were significant overall differences among the flight phases [Wilk's $\lambda = 0.000$, $F(38, 18) = 253.42$, $P < 0.001$, partial $\eta^2 = 0.998$]. Specifically, the ANOVA results and the post hoc pairwise comparison (Turkey HSD) results are shown in **Table I** for objective measurements and **Table II** for synthetic parameters, respectively.

Fixation duration, blink interval, and saccade rate did not show significant differences among flight phases. These parameters were integrated into the CGN dimension. The CGN did not show significant difference among flight phases either.

The deflection speed of the control column was highest in the approach landing phase and lowest in the cruise phase. The deflection speed of the control wheel and throttle were both lower in the cruise phase than in other phases, and they did not have significant differences between takeoff and approach landing. The deflection speed of the pedals was higher in approach landing than in other phases and it did not have significant differences between takeoff and cruise. These four parameters were integrated into the CTRL dimension, and it was highest in the approach landing phase and lowest in the cruise phase.

Respiration rate was higher in approach landing than in other phases. It did not have any significant differences between the takeoff and cruise phases. Respiration amplitude was lower in the approach landing phase than in the cruise phase. Heart rate and pupil diameter did not have significant differences among flight phases. These four parameters were integrated into the STRS dimension of workload. STRS was significantly higher in the approach landing phase than in other phases. It did not have significant differences between the takeoff and cruise phases.

The altitude deviation was significantly greater in the approach landing phase than in the takeoff phase. The cruise phase did not differ from other phases. The lateral deviation was higher in the cruise phase than in the takeoff phase. The approach landing phase did not differ from other phases. The airspeed deviation was lower in the cruise phase than in other phases. It did not have differences between the takeoff and approach landing phases. These three parameters were integrated into the PFM dimension. It was lower during cruise than in other phases and did not have significant differences between takeoff and approach landing.

The four workload dimensions were integrated into an overall workload assessment index. The overall workload assessment index showed significant differences among the three flight phases. It was higher in approach landing than in other phases. It did not have significant differences between takeoff and cruise. The absolute value of the Pearson product moment correlation coefficient between CGN and PFM, CTRL and PFM, CTRL and STRS, and CGN and STRS are shown in **Table III**.

In the takeoff phase, 20% of the participants had a significant correlation for CGN-PFM. In the cruise phase, none of the participants showed any significant correlations. In the approach landing phase, 40% of the participants showed significant correlation.

**Table I.** Differences of the Measurements Among Flight Phases.

| | | MEAN DIFFERENCE BETWEEN FLIGHT PHASES (TUKEY'S HSD) | | |
|---|---|---|---|---|
| PARAMETERS | $F$(2,27) | TAKEOFF-CRUISE | APPROACH-CRUISE | TAKEOFF-APPROACH |
| FD (ms) | 1.31 | 52.50 | −193.46 | 245.96 |
| BI (ms) | 2.08 | −29.82 | −47.29 | 17.47 |
| SR (saccades/min) | 0.04 | 5.51 | −1.29 | 6.79 |
| CS (°/min) | 28.95*** | 35.85* | 107.70*** | −71.84*** |
| WS (°/min) | 8.22** | 43.96** | 48.46** | −4.50 |
| PS (°/min) | 12.13*** | 2.22 | 12.75*** | −10.52** |
| TS (°/min) | 24.95*** | 69.21*** | 76.72*** | −7.51 |
| PD (mm) | 1.72 | 0.05 | −0.6 | 0.6 |
| HR (beats/min) | 1.04 | −3.57 | 6.27 | −9.84 |
| RR (breaths/min) | 7.84** | 2.40 | 8.33** | −5.93* |
| RA (V) | 4.77* | −0.006 | −0.011* | 0.005 |
| AD (ft) | 4.44* | −42.16 | 69.40 | −111.56* |
| LD (m) | 7.45** | −183.39** | −110.03 | −73.36 |
| ASD (kts) | 26.01*** | 20.08*** | 16.15*** | 3.93 |

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

FD, fixation duration; BI, blink interval; SR, saccade rate; CS, control column deflection speed; WS, control wheel deflection speed; PS, pedals deflection speed; TS, throttle deflection speed; PD, pupil diameter; HR, heart rate; RR, respiration rate; RA, respiration amplitude; AD, altitude deviation; LD, lateral deviation; ASD, airspeed deviation.

http://prime-pdf-watermark.prime-prod.pubfactory.com/ | 2025-02-05

**Table II.** Differences of the Synthetic Parameters Among Flight Phases.

| | | MEAN DIFFERENCE BETWEEN FLIGHT PHASES (TURKEY HSD) | | |
|---|---|---|---|---|
| **PARAMETERS** | ***F*(2,27)** | **TAKEOFF-CRUISE** | **APPROACH-CRUISE** | **TAKEOFF-APPROACH** |
| CGN | 0.78 | 0.08 | −0.08 | 0.17 |
| CTRL | 117.24*** | 0.32*** | 0.69*** | −0.38*** |
| STRS | 23.45*** | 0.07 | 1.31*** | −1.24*** |
| PFM | 8.51** | 0.38** | 0.42** | −0.04 |
| WI | 15.27*** | 0.24 | 0.79*** | −0.55** |

**$P < 0.01$; ***$P < 0.001$.

CGN, cognitive activity; CTRL, control activity; STRS, stress; PFM, flight performance; WI, overall workload assessment index.

For the CTRL-PFM correlation, in the takeoff phase, 10% of the participants had a significant correlation. In the cruise phase, the Pearson correlation coefficients did not exist because, during this phase, the aircraft was completely controlled by automatic systems. Therefore, pilots did not have any control activities. In the approach landing phase, 50% of the participants showed significant correlation.

For the CTRL-STRS correlation, in the takeoff phase, none of the participants had any significant correlations. In the cruise phase, the Pearson correlation coefficients did not exist since the aircraft was controlled by automatic systems. In the approach landing phase, 50% of the participants showed significant correlation.

For the CGN-STRS correlation, in the takeoff phase, 20% of the participants had a significant correlation. In the cruise phase, 10% of the participants showed a significant correlation. In the approach landing phase, 60% of the participants showed significant correlation.

## DISCUSSION

The NASA-TLX technology has already been validated with a number of studies in various fields. In this study, it served as criterion measure of workload. The experimental results indicated that generally the proposed integrated method and the NASA-TLX had similar results. Both of them showed that workload in the approach landing phase was significantly higher than in other phases. Neither of them showed differences between the takeoff and cruise phases. It is noteworthy that, compared to NASA-TLX, the proposed method is based on objective measuring which can continuously record the detailed information of the entire task process. It would not be affected by a participant's emotion or experience. It is not necessary to worry about participants' memory deviation.

Experimental results showed that the PCA-based integration method can inherit the sensitivity of the elementary objective measurements. Sometimes not all the measurements of a workload dimension had significant difference among flight phases. As shown in the STRS dimension, only respiration rate and respiration amplitude showed a difference. Heart rate and pupil diameter did not have significant differences. After integration, the STRS dimension can still distinguish the workload difference among flight phases.

Integrating multiple parameters into a workload dimension could also bring reliability to the assessment. In some dimensions, although all the measurements had significant differences among flight phases, they varied with different patterns. For example, in the CTRL dimension, the deflection speed of the control column differed between every two flight phases. The deflection speed of the control wheel and throttle did not have differences between the takeoff and approach landing phases. The deflection speed of the pedals did not have differences between the takeoff and cruise phases. If we only consider one of these parameters when evaluating a pilot's control activity, the result could be one-sided and inappropriate. Integrating all these parameters could provide an overall evaluation of a pilot's control activity.

Assessing pilot workload from multiple dimensions could reveal more details regarding a pilot's workload. For example, although the overall workload assessment index did not have any significant difference between the takeoff and cruise phases, the four workload dimensions indicated many differences between the two phases. The CTRL dimension showed that pilots had more control activities in the takeoff phase than in the cruise phase. The PFM dimension showed that the performance deviation was greater in the takeoff phase than in the cruise phase.

**Table III.** Correlations Between the Workload Dimensions in Different Flight Phases.

| | CGN-PFM | | | CTRL-PFM | | | CTRL-STRS | | | CGN-STRS | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SUBJECT** | **T** | **C** | **A** | **T** | **C** | **A** | **T** | **C** | **A** | **T** | **C** | **A** |
| 1 | 0.39 | 0.24 | 0.26 | 0.06 | / | 0.01 | 0.33 | / | 0.51* | 0.03 | 0.07 | 0.60* |
| 2 | 0.08 | 0.36 | 0.36 | 0.23 | / | 0.32 | 0.10 | / | 0.61** | 0.55* | 0.63* | 0.21 |
| 3 | 0.24 | 0.57 | 0.65** | 0.13 | / | 0.08 | 0.06 | / | 0.28 | 0.11 | 0.18 | 0.26 |
| 4 | 0.06 | 0.22 | 0.10 | 0.17 | / | 0.22 | 0.04 | / | 0.59* | 0.12 | 0.50 | 0.82*** |
| 5 | 0.55* | 0.55 | 0.41 | 0.09 | / | 0.72** | 0.31 | / | 0.38 | 0.17 | 0.40 | 0.49* |
| 6 | 0.43 | 0.05 | 0.05 | 0.27 | / | 0.78** | 0.22 | / | 0.83*** | 0.41 | 0.14 | 0.74** |
| 7 | 0.33 | 0.21 | 0.43 | 0.55* | / | 0.34 | 0.14 | / | 0.43 | 0.32 | 0.19 | 0.48* |
| 8 | 0.34 | 0.36 | 0.73*** | 0.05 | / | 0.53* | 0.31 | / | 0.15 | 0.01 | 0.20 | 0.10 |
| 9 | 0.08 | 0.57 | 0.46* | 0.08 | / | 0.63** | 0.06 | / | 0.22 | 0.09 | 0.27 | 0.14 |
| 10 | 0.57** | 0.46 | 0.63** | 0.39 | / | 0.75*** | 0.16 | / | 0.55* | 0.67** | 0.48 | 0.85*** |

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

CGN-PFM: cognitive activity-flight performance; CTRL-PFM: control activity-flight performance; CTRL-STRS: control activity-stress; CGN-STRS: cognitive activity-stress; T: takeoff; C: cruise; A: approach landing.

http://prime-pdf-watermark.prime-prod.pubfactory.com/ | 2025-02-05

It has to be noted that some workload dimensions did not show significant differences between certain phases. For example, STRS was significantly higher in the approach landing phase than in the other two flight phases. This result complies with the task difficulty settings (the task in the approach landing phase is more complex than the takeoff and cruise phases). Comparing the takeoff and cruise phases, the mean value of STRS is higher in the takeoff phase (see Table II); however, this difference did not show statistical significance. Actually, although there are more manual controls in the takeoff phase than in the cruise phase, these controls are standard step-by-step procedures which do not require too much mental or physical resource. For the CGN dimension, it did not have significant differences among the three flight phases. Little flight experience might be one reason for this result. In a previous study, it was found that more experienced pilots had lower saccade rates in emergency tasks than in normal tasks. However, for less experienced pilots, their saccade rate did not have significant differences between normal tasks and emergency tasks.[26] It shows that experienced pilots could flexibly adapt their visual perception strategy according to task demand. The subjects in this study had only 6 h of flight training, so it was hard for them to acquire such an ability.

The correlation between workload dimensions could bring additional diagnostic information. Sometimes the amount of a workload dimension did not differ between tasks. Experimental results indicated that the CGN dimension did not have differences between the takeoff and approach landing phases. However, it was found that a comparable amount of cognitive activities would have a different impact on pilots. The correlation analysis results in Table III illustrated that, in the takeoff phase, only 20% of the participants showed significant correlation between cognitive activity and stress. In the approach landing phase this number rose to 60%. Meanwhile, considering that the STRS dimension was higher in the approach landing phase, it revealed that the cognitive activity in the approach landing phase might cause more stress than that in the takeoff phase. There were also more pilots that showed significant CTRL-STRS correlation in the approach phase than in the takeoff phase (50% vs. 0%). Therefore, the control activities were also more stressful in the approach landing phase. More pilots showed significant CTRL-PFM correlation in the approach landing phase than in the takeoff phase (50% vs. 10%). It seemed that more pilots pay particular attention to flight performance when making control decisions in the approach landing phase than in the takeoff phase.

The correlations between workload dimensions also had important significance to cockpit design evaluation. For example, if a newly developed cockpit is to be compared against a reference model. If the CGN-PFM correlation in the newly developed cockpit is significantly higher than that in the reference model, we might deduce that the new design is more salient in catching a pilot's attention.

Some previously developed integrated workload assessment methods determined their model structures based on supervised machine learning. In implementation of these methods, model training is essential, and the model structure is completely based on the training dataset. Considering that a human operator introduces too much uncertainty,[21] it is extremely hard to acquire a satisfactory training dataset. In this study, the principal component analysis was applied for each trial independently. It used the information contained in the covariance matrix of the measured data and can adaptively determine the model structure. This method does not need an extra training process and is flexible to different types of task.

This study proposed a framework to assess pilot workload. Although in the implementation several objective indices were measured, these indices are not immutable. They can be extended and optimized. For example, according to Wickens' multiple resource theories,[27] besides visual activity, humans have mental resources for other kinds of activity such as auditory activity and tactual activity, etc. The validity of the workload dimensions could be improved by including other kinds of objective measurements.

Apart from the above characteristics of the proposed method, it has to be recognized that the study still has some limitations. Since this was a preliminary study, the experiment was conducted in a simulated environment (specifically in a fixed-base simulator) instead of actual flight. In this situation, control error and performance deviation will not really cause serious consequences. The subjects would be less sensitive to stress. Wilson et al. studied fighter pilots' physiological reactions under varying task demand in both actual flight and simulator conditions, and found that performing tasks in a simulator with six degrees of motion hardly influenced heart rate; however, significant heart rate changes were observed when the same tasks were performed in real flight.[31] Magnusson found that lack of sense of motion (changes of attitude and acceleration) would also influence physiological reactions. In his study, pilots' heart rate level in a fixed-base simulator was lower than in real flight.[14] This is also a potential reason for the insignificant difference of some parameters in our study. In a future study, it is necessary to validate the proposed method in actual flight with pilots of different levels and with various task scenarios.

In conclusion, this study developed a top-down approach to assess pilot workload. The approach describes workload from four dimensions: pilots' cognitive activity, control activity, stress, and flight performance. Each dimension is represented by integrating multiple objective parameters. The correlations between the workload dimensions were also considered. After being implemented in a simulated flight task, experimental results indicated that the proposed method is sensitive and reliable and can cope with the problems of single parameters such as limited sensitivity and vulnerability to noise. Describing workload from multiple dimensions provides detailed information about pilot aircraft interaction and the status of the man-machine system. Correlation between the workload dimensions could bring useful diagnostic information. Though the method still needs to be optimized in a future study, we think it is a promising tool and hope it can help evaluators and designers to improve flight safety.

http://prime-pdf-watermark.prime-prod.pubfactory.com/ | 2025-02-05

## ACKNOWLEDGMENTS

*Financial Disclosure Statement:* The authors have no competing interests to declare.

*Authors and affiliation:* Zhen Wang, Ph.D., M.S., Yanyu Lu, Ph.D., B.S., and Shan Fu, Ph.D., B.S., Department of Automation, Shanghai Jiao Tong University, Shanghai, China.

## REFERENCES

1. Annett J. Subjective rating scales: science or art? Ergonomics. 2002; 45(14):966–987.
2. Besson P, Bourdin C, Bringoux L, Dousset E, Maïano C, et al. Effectiveness of physiological and psychological features to estimate helicopter pilots' workload: a Bayesian network approach. IEEE Trans Intell Transp Syst. 2013; 14(4):1872–1881.
3. Borghini G, Astolfi L, Vecchiato G, Mattia D, Babiloni F. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neurosci Biobehav Rev. 2014; 44:58–75.
4. Cain B. A review of the mental workload literature. Toronto (Canada): Defence Research and Development Canada; 2007. Report No.: #RTO-TR-HFM-121-Part-II.
5. Charles RL, Nixon J. Measuring mental workload using physiological measures: a systematic review. Appl Ergon. 2019; 74:221–232.
6. Chueh T, Chen T, Lu H, Ju S, Tao T, Shaw J. Statistical prediction of emotional states by physiological signals with manova and machine learning. Int J Pattern Recognit Artif Intell. 2012; 26(04):1250008.
7. Harris D. Human performance on the flight deck. Aldershot: Ashgate; 2011.
8. Harris D, Gautrey J, Payne K, Bailey R. The Cranfield aircraft handling qualities rating scale: a multidimensional approach to the assessment of aircraft handling qualities. Aeronaut J. 2000; 104(1034):191–198.
9. Healey J, Picard R. Detecting stress in real-world driving tasks using physiological sensors. IEEE Trans Intell Transp Syst. 2005; 6(2):156–166.
10. Huberty C, Morris J. Multivariate analysis versus multiple univariate analyses. Psychol Bull. 1989; 105(2):302–308.
11. Kelly D, Efthymiou M. An analysis of human factors in fifty controlled flight into terrain aviation accidents from 2007 to 2017. J Safety Res. 2019; 69:155–165.
12. Lahtinen TMM, Koskelo JP, Laitinen T, Leino TK. Heart rate and performance during combat missions in a flight simulator. Aviat Space Environ Med. 2007; 78(4):387–391.
13. Lysaght RHill SDick APlamondon BLinton P. Operator workload: comprehensive review and evaluation of operator workload methodologies. Fort Bliss (TX, USA): United States Army Research Institute for the Behavioural and Social Sciences; 1989. Technical Report 851.
14. Magnusson S. Similarities and differences in psychophysiological reactions between simulated and real air-to-ground missions. Int J Aviat Psychol. 2002; 12(1):49–61.
15. Matthews G, Reinerman-Jones LE, Barber DJ, Abich J. The psychometrics of mental workload: multiple measures are sensitive but divergent. Hum Factors. 2015; 57(1):125–143.
16. Nakayama M, Takahashi K, Shimizu Y. The act of task difficulty and eye-movement frequency for the 'Oculo-motor indices'. ETRA 2002: Eye Tracking Research and Applications Symposium. New York: ACM; 2002:37–42.
17. Parasuraman R, Riley V. Humans and automation: use, misuse, disuse, abuse. Hum Factors. 1997; 39(2):230–253.
18. Pedrotti M, Mirzaei MA, Tedesco A, Chardonnet J-R, Merienne F, et al. Automatic stress classification with pupil diameter analysis. Int J Hum Comput Interact. 2014; 30(3):220–236.
19. Pretorius A, Cilliers PJ. Development of a mental workload index: a systems approach. Ergonomics. 2007; 50(9):1503–1515.
20. Ren P, Barreto A, Huang J, Gao Y, Ortega F, Adjouadi M. Off-line and on-line stress detection through processing of the pupil diameter signal. Ann Biomed Eng. 2014; 42(1):162–176.
21. Rong H, Tian J, Zhao T. Temporal uncertainty analysis of human errors based on interrelationships among multiple factors: a case of a Minuteman III missile accident. Appl Ergon. 2016; 52:196–206.
22. Shappell S, Detwiler C, Holcomb K, Hackworth C, Boquet A, Wiegmann DA. Human error and commercial aviation accidents: an analysis using the Human Factors Analysis and Classification System. Hum Factors. 2007; 49(2):227–242.
23. Stanton NA, Salmon PM, Walker GH, Baber C, Jenkins DP. Human factors methods: a practical guide for engineering and design. Aldershot: Ashgate; 2005.
24. Veltman JA, Gaillard AWK. Physiological workload reactions to increasing levels of task difficulty. Ergonomics. 1998; 41(5):656–669.
25. Wang Z, Fu S. An analysis of pilots' physiological reactions in different flight phases. In: Harris D, editor. International Conference on Engineering Psychology and Cognitive Ergonomics. Heraklion (Greece): Springer; 2014:94–103.
26. Wang Z, Zheng LX, Lu YY, Fu S. Physiological indices of pilots' abilities under varying task demands. Aerosp Med Hum Perform. 2016; 87(4): 375–381.
27. Wickens CD. Multiple resources and mental workload. Hum Factors. 2008; 50(3):449–455.
28. Wickens CD, Hollands JG, Banbury S, Parasuraman R. Engineering psychology and human performance, fourth ed. New York: Psychology Press; 2013.
29. Wierwille WW, Connor SA. Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator. Hum Factors. 1983; 25(1):1–16.
30. Wilson GF. An analysis of mental workload in pilots during flight using multiple psychophysiological measures. Int J Aviat Psychol. 2002; 12(1):3–18.
31. Wilson GF, Purvis B, Skelly J, Fullenkamp P, Davis I. Physiological data used to measure pilot workload in actual flight and simulator conditions. Proceedings of the Human Factors Society Annual Meeting. 1987; 31(7):779–783.
32. Wilson GF, Russell CA. Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. Hum Factors. 2003; 45(4):635–644.
33. Xi P, Law A, Goubran R, Shu C. Pilot workload prediction from ECG using deep convolutional neural networks. IEEE International Symposium on Medical Measurements and Applications (IEEE MeMeA); 26–28 June 2019; Istanbul, Turkey. Piscataway (NJ, USA): IEEE; 2019:1–6.
34. Yao Y-J, Chang Y-M, Xie X-P, Cao X-S, Sun X-Q, Wu Y-H. Heart rate and respiration responses to real traffic pattern flight. Appl Psychophysiol Biofeedback. 2008; 33(4):203–209.
35. Yu C-S, Wang EM-Y, Li W-C, Braithwaite G. Pilots' visual scan patterns and situation awareness in flight operations. Aviat Space Environ Med. 2014; 85(7):708–714.

http://prime-pdf-watermark.prime-prod.pubfactory.com/ | 2025-02-05