Streamlining NASA Task Load Index for Simulated Fighter Missions

Heikki Mansikka, Ph.D.[1,2]; Kai Virtanen, D.Sc. (Tech.)[1,2]

[1]Department of Military Technology, National Defence University, Helsinki, Finland; [2]Department of Mathematics and Systems Analysis, Aalto University, Helsinki, Finland

DOI: 10.3357/AMHP.6817.2026

ABSTRACT

INTRODUCTION: Mental workload (MWL) is a key factor in fighter pilot training, yet the widely used National Aeronautics and Space Administration Task Load Index (NASA-TLX) may contain dimensions that are less relevant when pilot training is conducted in a simulator. This study examined whether the number of dimensions required to assess MWL could be reduced without compromising the validity of the NASA-TLX.

METHODS: NASA-TLX data were collected from 716 high-fidelity F/A-18 simulator missions flown by 20 combat-ready pilots across several different exercises. Pilots provided weights for workload dimensions prior to missions and completed post-mission ratings of the dimensions. Analyses included descriptive statistics of weights and scores, principal component analysis, and rank-based correlations between overall workload estimates computed from reduced-dimension composites and those based on all six workload dimensions.

RESULTS: Mental demand received the highest weight, while physical demand was consistently lowest. Principal component analysis revealed two components: a cognitive–emotional factor (mental, temporal, frustration) and a performance–effort factor (own performance, effort). Physical demand showed low communality and negligible loadings. Correlation analyses demonstrated that a three-dimensional composite (mental, temporal, frustration) closely reproduced the overall workload determined using six dimensions, while adding own performance and effort further improved correspondence.

DISCUSSION: The findings indicate that the NASA-TLX can be streamlined for simulated fighter missions by omitting the physical demand dimension and, when brevity is essential, by using three dimensions (mental, temporal, frustration) without substantial loss of validity. This approach reduces pilot burden and simplifies administration, supporting efficient MWL assessment in training contexts.

INTRODUCTION

Mental workload (MWL) refers to the balance between task demands and the cognitive resources available to meet them. In fighter pilot training context, understanding MWL is essential, as an imbalance can impair tactical decision making directly or indirectly through its impact on situation awareness.[1] Traditionally, MWL has been assessed with physiological measures (e.g., heart rate, heart rate variability and subjective measures)[2]. These measures allow continuous, objective monitoring, but subjective methods have proven equally sensitive, valid, and reliable in the fighter pilot context – especially when used in flight simulators.[3]

Among subjective methods, the National Aeronautics and Space Administration Task Load Index (NASA-TLX) is widely used, e.g., in live–virtual–constructive (LVC) pilot training environments,[4,5] because it is non-intrusive, inexpensive, and quick to administer. Compared with unidimensional MWL questionnaires, such as Modified Cooper-Harper scale,[6] it captures multiple workload dimensions: mental demand (MD), physical demand (PD), temporal demand (TD), own performance (OP), frustration (FR), and effort (EF). In its weighted form, pilots first assign weights to each dimension based on their perceived contribution to overall workload (OWL) before providing post-mission ratings for each dimension. Then, the OWL index is calculated as a weighted average of the dimension scores. Due to several challenges in the original weighting procedure,[7] alternative procedures have been proposed.[8] The so-called Raw NASA-TLX omits the weighting step entirely, as the OWL index is calculated simply as the average of the dimension scores.

Although workload dimension scores are relatively quick to collect, the use of the NASA-TLX could be further streamlined in the context of fighter pilot training. Reducing pilots' score assessment burden becomes a critical requirement if scores are collected during simulated or live missions. Moreover, as the NASA-TLX was designed for general use across domains, not all dimensions may be equally relevant in simulated fighter training, where physical effort is minimal.

Efforts to reduce the number of dimensions have been made in other domains. In patient monitoring, Said et al. validated a modified raw NASA-TLX without the PD dimension, finding that the reduced scale maintained the validity of the original scale.[9] In surgical MWL assessment, Lowndes et al. reported that the physical demand and frustration dimensions contributed little to discrimination across task difficulty, suggesting that some dimensions may add noise rather than clarity.[10] In rehabilitation engineering, Parr et al. developed the PROS-TLX for prosthesis use; although they retained many MWL constructs, they noted that certain dimensions (e.g., time pressure) varied little across tasks.[11]

Collectively, the studies mentioned above suggest that the PD dimension often contributes minimal information in contexts with negligible physical effort, and that a reduced set of dimensions can perform as well as all the complete six-dimensional version of the NASA-TLX. The study presented in this short communication builds on the assumption that not all dimensions are needed in simulated fighter missions. Most modern fighter aircraft simulators are non-motion and the piloting task entails minimal physical workload. Therefore, the aim of this study is to examine whether some of the NASA-TLX dimensions could be ignored when evaluating fighter pilots' MWL during simulated fighter missions without substantially compromising the validity of the method. The analyses presented in the following sections confirm this expectation, demonstrating that streamlined composites of the workload dimensions reproduce OWL indices computed from the full six-dimensional version of the raw and weighted NASA-TLX with reasonable fidelity.

METHODS

Subjects

Twenty combat-ready F/A-18 fighter pilots served as subjects. The pilots had a mean flight experience of 548.5 hours (standard deviation = 218.2) on the F/A-18. All subjects were male. A written, informed consent was obtained from each subject.

Materials

The dataset consisted of NASA-TLX dimension scores collected from 716 simulator missions. These data were obtained during routine, mandatory simulator training in operational air combat scenarios, where NASA-TLX assessments were collected as part of regular training objectives. The data were analyzed retrospectively for the purposes of this study, and no additional interventions, experimental manipulations, or data collection involving human participants were conducted. The pilots in this study were undertaking regular, routine training, and completed the full NASA-TLX as per standard training practice. The reduced-dimension workload indices were derived retrospectively from this existing NASA-TLX dataset (incorporating those completed by the pilots in this study) by omitting the physical demand dimension during analysis. That is, no additional procedures or deviations from normal training were introduced for research purposes. Under the guidelines of Finnish National Board on Research Integrity, retrospective analysis of such routinely collected training data does not require prior ethics committee approval.

Procedure

Before flying the sorties, the pilots assigned weights for the workload dimensions using Swing method.[8] Each sortie was a flown in a high-fidelity, operational F/A-18 simulator. The sorties corresponded to pilots' routine simulator training in operational air combat scenarios. The flying tasks involved beyond-visual-range air combat missions. After each mission, the pilots completed the NASA-TLX questionnaire, providing scores for the workload dimensions on standard 0–100 scale. Later, OWL is referred to as *raw* when calculated without weighting of the dimensions and as *weighted* when calculated with weighting. The sample of operationally experienced pilots provided a robust basis for evaluating whether a reduced-dimension composite can produce a sufficiently accurate OWL estimate for use in pilot training.

Statistical Analysis

The purpose of analysis was to evaluate whether all six NASA-TLX dimensions are necessary to compute OWL in the dataset, or whether a reduced set of dimensions would provide an OWL estimate with acceptable accuracy. The assigned dimension weights were first analyzed. Then, a principal component analysis (PCA) was performed on the dimension scores to explore the latent structure of the six dimensions and to identify those that contribute minimally to the OWL estimate. Prior to the analysis of the scores, the dataset was screened for multivariate outliers using Mahalanobis distance.[12] Five cases out of 716 were identified as outliers and excluded, resulting in a final sample of N = 711.

Moreover, five versions of the weighted and raw OWL indices were created, ranging from a single-dimension measure (MD only) to a five-dimension mean (MD, TD, FR, OP, and EF). Rank-based correlations (Spearman's rho and Kendall's tau-b) were calculated between each composite and the six-dimensional OWL index in both its raw and weighted forms. This approach allowed us to quantify how well reduced-dimension composites reproduce the rank ordering of the OWL based on all six dimensions, providing empirical guidance for potential streamlining of NASA-TLX administration. The data were collected non-invasively during fighter pilots' routine training which they would have undertaken even in the absence of the experiment.

RESULTS

Descriptive statistics for the dimension weights and both raw and weighted OWL dimension scores are presented in **Tables I** and **II**. Both mean weights and mean dimension scores were highest for MD and EF, while PD receiving the lowest mean values, consistent with the low-physical-load nature of the simulated missions. The OWL indices averaged 43.2 (raw) and 47.0 (weighted).

Table I. Descriptive Statistics of Dimension Weights (N=20).

DIMENSION	MIN	MAX	M	SD
MD, weight	0.14	0.49	0.25	0.09
PD, weight	0.03	0.16	0.08	0.04
TD, weight	0.09	0.31	0.19	0.05
OP, weight	0.04	0.26	0.16	0.07
FR, weight	0.08	0.24	0.15	0.05
EF, weight	0.02	0.23	0.17	0.05

Min=minimum, Max=maximum, M=mean, SD=standard deviation.

Table II. Descriptive Statistics of Raw and Weighted Dimension Scores (N=711).

DIMENSION	MIN	MAX	M	SD
MD, raw score	10.00	90.00	51.82	16.96
PD, raw score	0.00	85.00	13.67	15.18
TD, raw score	0.00	95.00	49.49	20.66
OP, raw score	0.00	100.00	46.98	20.99
FR, raw score	10.00	90.00	54.70	17.25
EF, raw score	0.00	100.00	41.08	23.36
MD, weighted score	1.71	34.15	11.47	5.03
PD, weighted score	0.00	10.30	1.16	1.77
TD, weighted score	0.00	27.69	9.52	5.59
OP, weighted score	0.00	24.39	8.34	4.78
FR, weighted score	0.92	21.32	8.78	3.55
EF, weighted score	0.00	22.54	7.65	4.62

PCA of the dimension scores (see **Table III)** revealed a clear two-component structure explaining 65.1% of the variance in OWL. The first component was dominated by MD, TD, and FR, which together form a coherent ‘cognitive–emotional load’ factor. The second component was characterized by strong loadings for OP and EF, forming a distinct ‘performance–effort’ factor. PD showed low communality (.21) and weak loading on both components, indicating that it contributes only little to OWL. In practical terms, PD can be disregarded when interpreting latent workload dimensions.

Table III. Results of PCA on Dimension Scores (N=711).

DIMENSION	COMMUNALITY	LOADING (COMPONENT 1)	LOADING (COMPONENT 2)
MD	0.816	0.894	0.130
PD	0.212	0.460	0.020
TD	0.716	0.843	0.073
OP	0.696	0.061	0.832
FR	0.739	0.843	0.168
EF	0.726	0.083	0.848

Rank-based correlations between OWLs computed with different dimension composites and the raw and weighted OWL indices (see **Table IV**) confirmed that excluding PD does not substantially compromise fidelity. Spearman's rho (ρ) between the raw OWL index and progressively larger unweighted composites increased monotonically. Correlations with the weighted OWL index were slightly higher compared to those of the raw OWL. Kendall's tau-b (τ-b) values followed a similar pattern. As presented in **Table IV**, rank-based associations increased monotonically as dimensions were added. A three-dimensional composite (MD, TD, and FR) already correlated strongly with both the raw (ρ = 0.834, τ-b = 0.669) and weighted (ρ = 0.762, τ-b = 0.593) OWL index. Adding OP and EF increased correspondence further, resulting in high correlations with the raw (ρ = .982, τ-b = .936) and weighted (ρ = .968, τ-b = .874) OWL indices, implying only modest gains beyond the three-dimensional composite. Together, these results indicate that an unweighted average of MD, TD, FR, OP, and EF estimates OWL almost as accurately as using all six workload dimensions. Moreover, even a three-dimensional composite (MD, TD, and FR) closely approximates OWL estimated using all six dimensions.

Table IV. Spearman's ρ and Kendall's τ-b Correlations Between OWLs Computed with Different Dimension Composites and the Raw and Weighted OWL Indices (N = 711).

COMPOSITE	DIMENSIONS INCLUDED	SPEARMAN'S ρ WITH RAW OWL	SPEARMAN'S ρ WITH WEIGHTED OWL	KENDALL'S τ-b WITH RAW OWL	KENDALL'S τ-b WITH WEIGHTED OWL
A	MD	0.754	0.652	0.607	0.511
B	MD + TD	0.787	0.711	0.625	0.547
C	MD + TD + FR	0.831	0.762	0.669	0.593
D	MD + TD + FR + OP	0.935	0.891	0.808	0.735
E	MD + TD + FR + OP + EF	0.982	0.968	0.936	0.874

DISCUSSION

Simulators play an important role in fighter pilot training, and with modern single-seat fighter aircraft, their role is expected to further increase. When simulator resources are limited, even small time savings in MWL assessment will accumulate. Assuming equal time is required to score each NASA-TLX dimension, omitting one dimension reduces completion time by approximately 16%.

This study examined whether certain NASA-TLX dimensions could be excluded from MWL assessment without substantially compromising the validity of the OWL measure. As shown in **Table I**, the subjective weights indicated that subjects prioritized MD as the most influential factor, followed by TD, OP, and EF. FR was weighted somewhat lower, whereas the priority given to PD was low. These weights suggest that the pilots considered MWL primarily as a combination of cognitive demands, time pressure, and self-perceived performance and effort, while physical strain was largely irrelevant. The dimension scores summarized in **Table II** reflect the similar importance of dimensions.

PCA provided converging evidence supporting the assumption that the NASA-TLX can be reliably used in fighter pilots' simulator training without the need for all six workload dimensions.

As summarized in **Table III**, it revealed a two-component structure that explained 65.1% of the variance in OWL. The first component, dominated by MD, TD, and FR, represents a cognitive–emotional factor, while the second - driven by OP and EF - captures a performance–effort factor. PD showed low communality and negligible loadings, highlighting its low importance to OWL.

As presented in **Table IV**, PD contributed minimal unique information to OWL. Raw and weighted OWL values across different dimension composites were comparable, indicating that weighting provides little additional benefit for rank fidelity.

Overall, the results suggest that the number of workload dimensions in the NASA-TLX can be reduced in the context of fighter pilots' simulator training without remarkably compromising the method's validity. A three-dimensional composite (MD, TD, and FR) reproduced the OWL index computed from all six dimensions with reasonable fidelity, making it a strong candidate for operational use in simulator-based training. When greater diagnostic detail is required, a five-dimensional composite that adds OP and EF is recommended. These findings align with earlier results from healthcare, surgical, and rehabilitation contexts, where physical effort was also negligible. To summarize, this study demonstrated that the NASA-TLX can be tailored to the fighter pilot training context by excluding irrelevant workload dimensions without substantially compromising the measure's validity.

Future research should evaluate the convergent validity of a dimension-reduced NASA-TLX by comparing the OWL index derived from fewer than six workload dimensions with independent MWL measures, such as physiological indices (e.g., electroencephalography and electrocardiography) or performance-based metrics. Also, the generalizability of the findings beyond the simulated fighter context should be explored.

For applications where task demands resemble those of fighter pilots' simulator missions, it is recommended to focus on MD, TD, and FR as concise yet effective indicators of MWL. This streamlined approach allows for the evaluation of MWL while minimizing administration time and participant burden. For contexts that require more detailed analysis or higher diagnostic sensitivity, including OP and EF can provide additional insight into MWL, without the need to incorporate the PD dimension. By tailoring the NASA-TLX in this way, MWL monitoring can be more seamlessly integrated into routine simulator training, supporting targeted interventions, optimizing training efficiency, and ultimately enhancing both safety and operational performance. These findings may also inform MWL assessment in other high-demand operational settings where rapid, reliable evaluation is crucial.

REFERENCES

1. Mansikka H, Virtanen K, Harris D. Dissociation between mental workload, performance, and task awareness in pilots of high performance aircraft. IEEE Trans Hum Mach Syst. 2019 Feb; 49(1):1-9.
2. Mansikka H, Virtanen K, Harris D, Simola P. Fighter pilots' heart rate, heart rate variation and performance during an instrument flight rules proficiency test. Appl Ergon. 2016 Sep; 56:213-9.
3. Mansikka H, Harris D, Virtanen K. Comparison of NASA-TLX scale, modified Cooper-Harper scale and mean inter-beat interval as measures of pilot mental workload during simulated flight tasks. Ergonomics. 2018 Apr; 62(2):246-54.
4. Mansikka H, Virtanen K, Harris D, Salomäki J. Live–virtual–constructive simulation for testing and evaluation of air combat tactics, techniques, and procedures, Part 1: assessment framework. JDMS. 2012 Nov; 18(4):285-93.
5. Mansikka H, Virtanen K, Harris D, Salomäki J. Live–virtual–constructive simulation for testing and evaluation of air combat tactics, techniques, and procedures, Part 2: demonstration of the framework. JDMS. 2012 Nov; 18(4):295-308.
6. Wierwille WW, Casali JG. A validated rating scale for global mental workload measurement applications. Proc Hum Factors Ergon Soc Annu Meet. 1983 Oct; 27(2):129–33.
7. Hart SG, Staveland LE. Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock PA, Meshkati, N, editors. Human mental workload. New York (NY): North Holland; 1988. p. 139-83. (Advances in psychology; 52).

8. Virtanen K, Mansikka H, Kontio H, Harris D. Weight watchers: NASA TLX weights revisited. Theor Issues Ergon Sci. 2021 Nov; 23(6):725-48.
9. Said S, Gozdzik M, Roche TR, Braun J, Rössler J, Kaserer A, et al. Validation of the raw National Aeronautics and Space Administration Task Load Index (NASA-TLX) questionnaire to assess perceived workload in patient monitoring tasks: pooled analysis study using mixed models. J Med Internet Res. 2020 Sep;22(9):e19472. doi: 10.2196/19472. PubMed PMID: 32780712; PubMed Central PMCID: PMC7506540.
10. Lowndes B, Forsyth K, Blocker R, Dean P, Truty M, Heller S, et al. NASA-TLX assessment of surgeon workload variation across specialties. Ann Surg. 2020 Apr; 271(4):686-92.
11. Parr J, Galpin A, Uiga L, Marshall B, Wright D, Franklin Z, et al. A tool for measuring mental workload during prosthesis use: The prosthesis task load index (PROS-TLX). PLoS One. 2023 May; 18(5):1-15.
12. Leys C, Klein O, Dominicy Y, Ley C. Detecting multivariate outliers: use a robust variant of the Mahalanobis distance. J Exp Soc Psychol. 2018 Jan; 74:150-6.