

Comparison of Mental Workload with N-Back Test: A New Design for NASA-Task Load Index Questionnaire

Mahdi Malakoutikhah¹, Reza Kazemi², Hadiseh Rabiei³, Moslem Alimohammadlou⁴, Asma Zare⁵, Soheil Hassanipour⁶

¹Department of Occupational Health, Kashan University of Medical Sciences, Kashan, ²Department of Ergonomics, School of Health, Shiraz University of Medical Sciences, ⁴Department of Industrial Management, Faculty of Economic, Management and Social Science, Shiraz University, ⁵Department of Industrial Management, Faculty of Economic, Management and Social Science, Shiraz, ³Department of Occupational Health Engineering, School of Health, Shahid Beheshti University of Medical Sciences, Tehran, ⁶Cardiovascular Diseases Research Center, Department of Cardiology, Heshmat Hospital, School of Medicine, Guilan University of Medical Sciences, Rasht, Iran

ORCID:

Moslem Alimohammadlou 0000-0002-9463-7285

Abstract

Aims: A most widely used tool for measuring the workload is the NASA-task load index (TLX) questionnaire, in which various studies have reported numerous problems. The present study aimed to improve the NSAS-TLX mental workload questionnaire using the fuzzy linguistic variables instead of the virtual rating scale, and multicriteria decision-making Fuzzy Best-Worst method (FBWM) instead of pair-wise comparison. **Materials and Methods:** This cross-sectional study was carried out among students of Shiraz University of Medical Sciences. In order to compare the traditional NASA-TLX and FBWM NASA-TLX questionnaire, participants performed a standard N-Back task with three low, medium, and high workload, subsequently, they completed the two questionnaires. Finally, results were examined using the researcher-made software and SPSS 16. **Results:** With regard to the reaction time to response the N-Back test, the mean of correct responses were 107.43 and 85.86 responses at levels 1 and 3, respectively. The mean score of the subscales and the final score in the two questionnaires at different levels of the N-Back test were independent as follows: the two questionnaires had a significant difference on mental demand at level 3 with a mean (standard deviation [SD]) of 18.09 (6.39) in the FBWM-NASA-TLX questionnaire and mean (SD) of 22.64 (8.15) in the NASA-TLX questionnaire ($P < 0.05$). **Conclusion:** In this study, the FBWM-NASA-TLX questionnaire was designed and studied with regard to the problems and limitations of the NASA-TLX questionnaire. Results of this study showed that the FBWM-NASA-TLX questionnaire can estimate more realistic scores and decisions of workload in the studied task.

Keywords: Ergonomics, fuzzy best-worst method, mental workload, NASA-task load index

INTRODUCTION

Workload is a term applied to describe the amount of cognitive and physical resources used to perform a task.^[1] According to Hart and Staveland, workload is defined as a presumptive structure representing the costs provided by the operator to achieve a certain level of performance.^[2,3] Nowadays, the trend to examine the mental workload (MWL) among scholars has increased due to change in the nature of tasks from physical to cognitive demands.^[4] As a result, assessment of MWL is essential as much as the physical workload. MWL, especially in studying and developing the human-machine interactions,

is important to achieve appropriate levels of satisfaction, comfort, safety, and efficiency at the workplace, which is considered as one of the main goals of ergonomics.^[5] Hence, MWL has become one of the most commonly used concepts in ergonomics study and practice.^[6-11]

Although there are many different subjective and objective methods to evaluate MWL, the NASA-task load index (TLX) questionnaire, first presented by the United States National

Address for correspondence: Dr. Moslem Alimohammadlou, Department of Industrial Management, Faculty of Economic, Management and Social Science, Shiraz University, Shiraz, Iran. E-mail: [msلمامل@shirazu.ac.ir](mailto:mسلمامل@shirazu.ac.ir)

Received: 05-Nov-2020

Revised: 05-Nov-2020

Accepted: 07-Dec-2020

Published: 31-Mar-2021

Access this article online

Quick Response Code:



Website:
<http://iahs.kaums.ac.ir>

DOI:
10.4103/iahs.iahs_126_20

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Malakoutikhah M, Kazemi R, Rabiei H, Alimohammadlou M, Zare A, Hassanipour S. Comparison of mental workload with N-Back test: A new design for NASA-task load index questionnaire. *Int Arch Health Sci* 2021;8:7-13.

Archive of SID

Aeronautics and Space Administration, has high validity and acceptance due to its multifaceted features and has been used in previous studies.^[12-14] This questionnaire consists of two sections, (1) rating the intensity of each subscale in the range between 0 and 100 and (2) comparing the six subscales in two by two manner.^[15,16] Different studies have presented some problems in the results obtained from the NASA-TLX index.^[17,18] In the first section of the questionnaire, a visual rating scale is used, where the participant must mark specific lines. The experience shows that people are interested in marking between the lines; hence, the researcher cannot achieve the specified number.^[19]

The most important problem in this questionnaire is its pair-wise weighting. Pair-wise weighting can easily be calculated, but some methodological and practical problems might be risen, especially in the actual working environments. First, the weighted mean is based on mathematical assumptions that are not usually approved.^[16] The second issue is the fact that it cannot refer to variable interactions of workload to correctly represent the integration or effectiveness of a subscale.^[17] Third, the association of pair-wise weighting with raw NASA-TLX has also been questioned by others.^[16-18] Furthermore, in the pair-wise weighting of this index, one should only choose one of the two options, while both subscales might be equally important in a task. In addition, the importance level of each subscale cannot be specified by choosing only one option.^[20]

With regard to the problems of NASA-TLX questionnaire; the aim of this study was to improve the NSAS-TLX questionnaire using the Fuzzy Best-Worst method (FBWM). Therefore, this study will present a new instrument to measure the workload using a general change in the NASA-TLX questionnaire.

METHODS

Procedure

This cross-sectional study was carried out at Shiraz University of Medical Sciences. In order to evaluate the conventional NASA-TLX and proposed FBWM NASA-TLX questionnaires, the participants performed N-Back standard task at three low, moderate, and high workloads. After each level, the perceived MWL was assessed using the conventional and proposed questionnaires. At first, the participants were trained for 5 min to perform the test and then, they randomly selected a task from three levels of N-Back task and performed it. After completing each task, the questionnaires were immediately given to the participants and they were asked to accurately express their perceived MWL proportional to the performed task. After completing each task and questionnaire, participants rested for 15 min, so that mental exhaustion would not affect the performance of the individual for the next task. Written informed consent was obtained.

Ethics

This ethics of the study was approved by Shiraz University of Medical Sciences (SUMS) with ethic code No. IR.SUMS.REC.1397.942.

Study design

Selection and description of participants

Participants were selected from the students of SUMS randomly. Inclusion criteria were having mental and physical health, nondrug users, nonuse of nerve stimulants drugs, and adequate sleep before the study. The demographic characteristics of participants are presented in Table 1.

Technical information

N-back test

This test is considered as one of the most widely used instruments for measuring the active memory and is a cognitive function assessment task related to executive actions.^[21] The participant should check whether the current stimulus is similar to the previous n-step stimulus or not. The difficulty level of this test depends on the comparison of the stimulus with the n-step stimulus, so that if it is 2-back, the participant should compare the current stimulus with the two previous step stimulus.^[22]

NASA-task load index questionnaire

It is a six-dimensional scale for estimating workload. The index originally consisted of two sections. The total workload of an activity is divided into six subscales, including Mental demand (MD), Performance (PE), Effort (EF), Frustration (FR), Temporal demand (TD), and Physical demand (PD).

FBWM-NASA-TLX questionnaire: like the NASA-TLX, the questionnaire designed for this study consists of two parts, estimating the intensity of each subscale in a fuzzy function and weighing each subscale using the FBWM.

Section 1

The present study uses the fuzzy linguistic variables (membership functions) instead of the numerical and visual rating scale.^[23] In other words, the participants first had to choose the intensity of each subscale by a term such as very low (0, 0, 25), moderate (0, 25, 50), high (25, 50, 75), and very high (75, 100, 100).

Section 2

FBWM was proposed by Guo and Zhao in 2017.^[24] In the FBWM method, instead of pair-wise comparisons of the variables, comparisons and conclusion are carried out in four steps: (a) selecting the best and worst criteria; (b) comparing

Table 1: Demographic characteristics of participants

Characteristics	n (%)	Mean (SD)
Gender		
Male	10 (33.3)	-
Female	20 (66.7)	
Education		
B.Sc.	17 (56.7)	
M.Sc.	5 (16.7)	
Ph.D.	8 (26.7)	
Total	30 (100)	
Age	-	24.9 (4.79)

SD: Standard deviation

Archive of SID

the degree and intensity of importance of the best criteria toward other criteria with linguistic variables; (c) comparing the degree and intensity of importance of all criteria toward the worst criterion with linguistic variables [Table 2]; and d) calculating the final weight of the criteria. In this way, frequent comparisons and their large number are prevented, and the decision maker can simply make a better and more accurate decision.^[25] Then, Eqs. (1) to (4) are used to weigh the criteria.

Eq. (1): The function of comparing the best criterion with other criteria

$$\tilde{A}_B = (a_{B1}, a_{B2}, a_{B3}, \dots, a_{Bn}) \quad (1)$$

Where \tilde{A}_B is the function of comparing the best criterion with other criteria, a_{Bj} represents the chosen linguistic variable to represent the importance degree of the best criterion toward the criterion j .

Eq. (2): The function of comparing other criteria toward the worst criterion.

$$\tilde{A}_W = (a_{1W}, a_{2W}, a_{3W}, \dots, a_{nW}) \quad (2)$$

Where \tilde{A}_W is the function of comparing other criteria with the worst criterion, represents the linguistic variable to represent the importance degree of the other criterion i to the worst criterion.

Eq.(3) is the final equation for weighting the criteria in FBWM.

$$\begin{aligned} & \text{m in } \max_j \left\{ \left| \frac{\tilde{W}_B}{\tilde{W}_j} - \tilde{a}_{Bj} \right|, \left| \frac{\tilde{W}_j}{\tilde{W}_W} - \tilde{a}_{jW} \right| \right\} \\ & \text{s.t. } \begin{cases} \sum_{j=1}^n R(\tilde{W}_j) = 1 \\ l_j^w \leq m_j^w \leq u_j^w \\ l_j^w \geq 0 \\ j = 1, 2, \dots, n \end{cases} \end{aligned} \quad (3)$$

Where

$$\tilde{W}_B = (l_B^w, m_B^w, u_B^w), \tilde{W}_j = (l_j^w, m_j^w, u_j^w), \tilde{W}_j = (l_j^w, m_j^w, u_j^w), \tilde{a}_{Bj} = (l_{Bj}, m_{Bj}, u_{Bj})$$

$$\text{and } \tilde{a}_{jW} = (l_{jW}, m_{jW}, u_{jW})$$

By placing triangular fuzzy numbers in the equation above, Eq. (4) is formed.

Table 2: Transformation rules of linguistic variables of decision-makers

Linguistic terms	Membership function
Equally importance	1, 1, 1
Weakly important	0.67, 1, 1.5
Fairly important	1.5, 2, 2.5
Very important	2.5, 3, 3.5
Absolutely important	3.5, 4, 4.5

$$\begin{aligned} & \min \quad k \left\{ \left| \frac{\tilde{W}_B}{\tilde{W}_j} - \tilde{a}_{Bj} \right|, \left| \frac{\tilde{W}_j}{\tilde{W}_W} - \tilde{a}_{jW} \right| \right\} \\ & \text{s.t. } \begin{cases} \left| \frac{(l_B^w, m_B^w, u_B^w)}{(l_j^w, m_j^w, u_j^w)} - (l_{Bj}, m_{Bj}, u_{Bj}) \right| \leq k \\ \left| \frac{(l_j^w, m_j^w, u_j^w)}{(l_W^w, m_W^w, u_W^w)} - (l_{jW}, m_{jW}, u_{jW}) \right| \leq k \\ \sum_{j=1}^n R(\tilde{W}_j) = 1 \\ l_j^w \leq m_j^w \leq u_j^w \\ l_j^w \geq 0 \\ j = 1, 2, \dots, n \end{cases} \end{aligned}$$

Section 3: Calculating the final score of each subscale and the total score.

Finally, after completing both sections of the questionnaire by multiplying the two fuzzy numbers, the first section of the questionnaire was multiplied by the second part (Eq. 5), and finally, the final score of each subscale was achieved by difuzzification of the number obtained from the fuzzy multiplication of the two-section multiplication, using the Eq. (6).

$$\begin{aligned} & A = (l_1, m_1, u_1) \\ & B = (l_2, m_2, u_2) \\ & A \otimes B = (l_1 \otimes l_2, m_1 \otimes m_2, u_1 \otimes u_2) = (l_3, m_3, u_3) \end{aligned} \quad (5)$$

Where, A and B are the first and second sets, respectively, with three lower, medium, and upper bounds.

$$R(a_i) = \frac{l_i + 4m_i + u_i}{6} \quad (6)$$

Where, $R(a_i)$ is the difuzzified number obtained from the product of the two sections of the FBWM-NASA-TLX questionnaire. In fact, this is the final score of each subscale. The total score of the questionnaire is also obtained using the algebraic summation of the final score of each subscale.

Statistics

Microsoft Excel software used for calculating the FBWM-NASA-TLX questionnaire based on the above-mentioned equations and independent t -test was used by SPSS from IBM, North Castle, New York, U.S. (significance level $P < 0.05$).

RESULTS

Calculating the scores for subscales of the FBWM-NASA-TLX questionnaire (an example).

One of the questionnaires completed by the participants was as follows. After the N-Back test at level three in the first part

Archive of SID

of the questionnaire, the participant had selected the high level of linguistic variable for mental demand subscale, very high linguistic variable for performance, medium linguistic variable for effort, high linguistic variable for frustration, high linguistic variable for temporal demand, and very low linguistic variable for physical demand were chosen.

After calculating the Eq. (4), the weight of each subscale was obtained, and in this example, the fuzzy weight of mental demand subscale was (0.23, 0.23, 0.29), for performance (0.27, 0.27, 0.30), effort (0.06, 0.06, 0.07), frustration (0.13, 0.13, 0.17), temporal demand (0.20, 0.21, 0.37), and physical load (0.06, 0.06, 0.07). Fuzzy weight and final defuzzified weight of the mental demand subscale are (17.22, 22.96, 29.26) and 23.04, performance (20.35, 27.13, 30.26) and 26.52, effort (1.57, 3.13, 5.37) and 3.24, frustration (6.26, 9.39, 17) and 10.09, temporal demand (9.90, 15.65, 37) and 18.33, physical demand (0, 0, 1.79) and 0.30, respectively. Eventually, the obtained workload for this person was 81.52.

min k

$$\left\{ \begin{array}{l} l_1 - 1u_2 \leq ku_2; l_1 - 1u_2 \geq -ku_2 \\ m_1 - 1m_2 \leq km_2; m_1 - 1m_2 \geq -km_2 \\ u_1 - 1l_2 \leq kl_2; u_1 - 1l_2 \geq -kl_2 \\ l_1 - 3.5u_3 \leq ku_3; l_1 - 3.5u_3 \geq -ku_3 \\ m_1 - 4m_3 \leq km_3; m_1 - 4m_3 \geq -km_3 \\ u_1 - 4.5l_3 \leq kl_3; u_1 - 4.5l_3 \geq -kl_3 \\ l_1 - 1.5u_4 \leq ku_4; l_1 - 1.5u_4 \geq -ku_4 \\ m_1 - 2m_4 \leq km_4; m_1 - 2m_4 \geq -km_4 \\ u_1 - 2.5l_4 \leq kl_4; u_1 - 2.5l_4 \geq -kl_4 \\ l_1 - 0.67u_5 \leq ku_5; l_1 - 0.67u_5 \geq -ku_5 \\ m_1 - 1m_5 \leq km_5; m_1 - 1m_5 \geq -km_5 \\ u_1 - 1.5l_5 \leq kl_5; u_1 - 1.5l_5 \geq -kl_5 \\ l_1 - 3.5u_6 \leq ku_6; l_1 - 3.5u_6 \geq -ku_6 \\ m_1 - 4m_6 \leq km_6; m_1 - 4m_6 \geq -km_6 \\ u_1 - 4.5l_6 \leq kl_6; u_1 - 4.5l_6 \geq -kl_6 \\ l_2 - 3.5u_6 \leq ku_6; l_2 - 3.5u_6 \geq -ku_6 \\ m_2 - 4m_6 \leq km_6; m_2 - 4m_6 \geq -km_6 \\ u_2 - 4.5l_6 \leq kl_6; u_2 - 4.5l_6 \geq -kl_6 \\ l_3 - 1u_6 \leq ku_6; l_3 - 1u_6 \geq -ku_6 \\ m_3 - 1m_6 \leq km_6; m_3 - 1m_6 \geq -km_6 \\ u_3 - 1l_6 \leq kl_6; u_3 - 1l_6 \geq -kl_6 \\ l_4 - 1.5u_6 \leq ku_6; l_4 - 1.5u_6 \geq -ku_6 \\ m_4 - 2m_6 \leq km_6; m_4 - 2m_6 \geq -km_6 \end{array} \right.$$

$$\left\{ \begin{array}{l} u_4 - 2.5l_6 \leq kl_6; u_4 - 2.5l_6 \geq -kl_6 \\ l_5 - 2.5u_6 \leq ku_6; l_5 - 2.5u_6 \geq -ku_6 \\ m_5 - 3m_6 \leq km_6; m_5 - 3m_6 \geq -km_6 \\ u_5 - 3.5l_6 \leq kl_6; u_5 - 3.5l_6 \geq -kl_6 \\ \frac{1}{6} \times l_1 + 4 \times m_1 + \frac{1}{6} \times u_1 + \frac{1}{6} \times l_2 + 4 \times m_2 + \\ \frac{1}{6} \times u_2 + \frac{1}{6} \times l_3 + 4 \times m_3 + \frac{1}{6} \times u_3 \\ + \frac{1}{6} \times l_4 + 4 \times m_4 + \frac{1}{6} \times u_4 + \frac{1}{6} \times l_5 + 4 \times m_5 + \\ \frac{1}{6} \times u_5 + \frac{1}{6} \times l_6 + 4 \times m_6 + \frac{1}{6} \times u_6 = 1 \\ l_1 \leq m_1 \leq u_1 \\ l_2 \leq m_2 \leq u_2 \\ l_3 \leq m_3 \leq u_3 \\ l_4 \leq m_4 \leq u_4 \\ l_5 \leq m_5 \leq u_5 \\ l_6 \leq m_6 \leq u_6 \\ l_1 > 0; l_2 > 0; l_3 > 0; l_4 > 0; l_5 > 0; l_6 > 0 \\ k \geq 0 \end{array} \right. \tag{7}$$

Mental Demand:

$$\begin{aligned} \text{Part 1} \otimes \text{Part 2} &= (l_1 \otimes l_2, m_1 \otimes m_2, u_1 \otimes u_2) \\ &= (75 \times 0.23, 100 \times 0.23, 100 \times 0.29) \\ &= (17.22, 22.96, 29.22) \end{aligned} \tag{8}$$

Performance:

$$\begin{aligned} \text{Part 1} \otimes \text{Part 2} &= (l_1 \otimes l_2, m_1 \otimes m_2, u_1 \otimes u_2) \\ &= (75 \times 0.27, 100 \times 0.27, 100 \times 0.30) \\ &= (20.35, 27.13, 30.26) \end{aligned}$$

Effort:

$$\begin{aligned} \text{Part 1} \otimes \text{Part 2} &= (l_1 \otimes l_2, m_1 \otimes m_2, u_1 \otimes u_2) \\ &= (25 \times 0.06, 50 \times 0.06, 75 \times 0.07) \\ &= (1.57, 3.13, 5.37) \end{aligned}$$

Frustration :

$$\begin{aligned} \text{Part}1 \otimes \text{Part}2 &= (l_1 \otimes l_2, m_1 \otimes m_2, u_1 \otimes u_2) \\ &= (50 \times 0.13, 75 \times 0.13, 100 \times 0.17) \\ &= (6.26, 9.39, 17) \end{aligned}$$

Temporal Demand:

$$\begin{aligned} \text{Part 1} \otimes \text{Part 2} &= (l_1 \otimes l_2, m_1 \otimes m_2, u_1 \otimes u_2) \\ &= (50 \times 0.20, 75 \times 0.21, 100 \times 0.37) \\ &= (9.90, 15.65, 37) \end{aligned}$$

Archive of SID

Physical Demand:

$$\begin{aligned} \text{Part 1} \otimes \text{Part 2} &= (l_1 \otimes l_2, m_1 \otimes m_2, u_1 \otimes u_2) \\ &= (0 \times 0.06, 0 \times 0.06, 25 \times 0.07) \\ &= (0, 0, 1.79) \end{aligned}$$

$$\begin{aligned} R(\text{Mental Demand}) &= \frac{l_i + 4m_i + u_i}{6} \\ &= \frac{(17.22 + (4 \times 22.96) + (29.22))}{6} = 23.04 \end{aligned}$$

$$\begin{aligned} R(\text{Performance}) &= \frac{l_i + 4m_i + u_i}{6} \\ &= \frac{(20.35 + (4 \times 27.13) + (30.26))}{6} = 26.52 \end{aligned}$$

$$\begin{aligned} R(\text{Effort}) &= \frac{l_i + 4m_i + u_i}{6} \\ &= \frac{(1.57 + (4 \times 3.13) + (5.37))}{6} = 3.24 \end{aligned}$$

$$\begin{aligned} R(\text{Frustration}) &= \frac{l_i + 4m_i + u_i}{6} \\ &= \frac{(6.26 + (4 \times 9.39) + (17))}{6} = 10.09 \end{aligned}$$

$$\begin{aligned} R(\text{Temporal Demand}) &= \frac{l_i + 4m_i + u_i}{6} \\ &= \frac{(9.90 + (4 \times 15.65) + (37))}{6} = 18.33 \end{aligned}$$

$$\begin{aligned} R(\text{Physical Demand}) &= \frac{l_i + 4m_i + u_i}{6} \\ &= \frac{(0 + (4 \times 0) + (1.79))}{6} = 0.30 \end{aligned}$$

Comparing the NASA-task load index and Fuzzy Best-Worst method-NASA-task load index questionnaires

The mean scores of the subscales and the final score of the two questionnaires were evaluated separately in different levels of the N-Back test [Table 3]. As it can be seen, there was a significant difference between the two questionnaires on mental demands at the level three and physical load at levels two and three of the tests ($P < 0.05$).

Comparing different levels of the N-Back test and each questionnaire was performed separately by *t*-test [Tables 4 and 5]. This test showed that in the NASA-TLX questionnaire, there is a significant difference among the different levels of the N-Back test in subscales, including mental load, effort, frustration, and final score. In FBWM-NASA-TLX questionnaire, there is a significant difference between subscales, including mental load, performance, effort, and final score ($P < 0.05$).

DISCUSSION

The present study aimed to provide a novel instrument for measuring MWL with a fundamental change in the NASA-TLX questionnaire. As the results showed, at the most difficult level

of the N-Back test (level 3), the two questionnaires were significantly different in mental demand and physical load sub-scales at levels 2 and 3.

The N-Back test used in this study was a complete mental task, and by increasing the level of test and difficulty of doing it, the participants' mental demand increases.^[26] Hence, it can be stated that the most important subscale of this study is the same as mental demand and the significant difference between the two questionnaires in this subscale shows the superiority of the developed instrument. There was a significant difference between the two questionnaires in the mental load subscale, and the mean scores of this subscale were lower in the FBWM-NASA-TLX questionnaire, which is due to the different way of rating in the first part and weighting in the second part, and the final calculations of the questionnaire.

Furthermore, in pair-wise comparison, one has to select an option and he cannot express the same importance and equality between the two variables, which leads to unwanted option to receive

Table 3: Comparison of total scores of subscales between NASA-task load index and fuzzy best-worst method-NASA-task load index among N-Back levels

N-Back levels	Subscales	Mean (SD)		P*
		NASA-TLX	FBWM-NASA-TLX	
Level 1	Mental demand	11.23 (6.60)	9.98 (6.16)	0.455
	Performance	12.54 (7.18)	12.36 (6.59)	0.918
	Effort	7.05 (4.32)	6.63 (4.31)	0.706
	Frustration	3.56 (5.42)	4.11 (6.51)	0.727
	Temporal demand	11.42 (9.36)	9.74 (7.58)	0.444
	Physical demand	0.38 (0.65)	0.71 (0.81)	0.082
	Total	46.20 (16.00)	43.55 (17.51)	0.543
Level 2	Mental demand	16.16 (7.42)	14.04 (6.99)	0.258
	Performance	9.39 (5.95)	8.39 (4.93)	0.485
	Effort	9.16 (5.57)	8.45 (4.48)	0.588
	Frustration	6.78 (8.34)	5.55 (5.16)	0.494
	Temporal demand	13.58 (9.39)	12.35 (8.09)	0.588
	Physical demand	0.32 (0.73)	1.41 (2.37)	0.019*
	Total	55.42 (14.65)	50.21 (14.30)	0.169
Level 3	Mental demand	22.64 (8.15)	18.09 (6.39)	0.019*
	Performance	8.70 (7.41)	7.96 (6.72)	0.687
	Effort	10.83 (5.64)	10.02 (4.61)	0.544
	Frustration	8.96 (9.29)	7.93 (6.79)	0.626
	Temporal demand	13.30 (10.56)	13.60 (9.50)	0.909
	Physical demand	0.61 (1.57)	1.81 (2.10)	0.015*
	Total	65.06 (18.54)	59.42 (19.92)	0.262

*Independent *t*-test, **Statistically significant. SD: Standard deviation, FBWM: Fuzzy best-worst method, TLX: Task load index, NASA: National Aeronautics and Space Administration

Archive of SID

Table 4: The scores of subscales and total scores in NASA-task load index questionnaire by the different levels of the N-Back test

Subscales	N-Back levels, mean (SD)			P*
	Level 1	Level 2	Level 3	
Mental demand	11.23 (6.60)	16.17 (7.42)	22.64 (8.15)	<0.001**
Performance	12.54 (7.18)	9.39 (5.96)	8.70 (7.42)	0.075
Effort	7.05 (4.31)	9.17 (5.57)	10.83 (5.64)	0.023**
Frustration	3.57 (5.42)	6.79 (8.34)	8.97 (9.29)	0.032**
Temporal demand	11.42 (9.24)	13.59 (9.39)	13.30 (10.57)	0.648
Physical demand	0.38 (0.65)	0.32 (0.73)	0.61 (1.57)	0.543
Total	46.20 (15.99)	55.42 (14.65)	65.06 (18.54)	<0.001**

*Independent *t*-test, **Statistically significant. SD: Standard deviation**Table 5: The scores of subscales and total scores in fuzzy best-worst method-NASA-task load index questionnaire by the different levels of the N-Back test**

Subscales	N-Back levels, mean (SD)			P*
	Level 1	Level 2	Level 3	
Mental demand	9.99 (6.19)	14.03 (6.99)	18.09 (6.39)	<0.001**
Performance	12.36 (6.59)	8.39 (4.93)	7.96 (6.72)	0.012**
Effort	6.36 (4.31)	8.45 (4.48)	10.02 (4.61)	0.016**
Frustration	4.10 (6.51)	5.55 (5.16)	7.93 (6.79)	0.059
Temporal demand	9.74 (5.58)	12.35 (8.09)	13.60 (9.50)	0.201
Physical demand	0.72 (0.82)	1.41 (2.37)	1.81 (2.11)	0.080
Total	43.55 (17.51)	50.21 (14.30)	59.42 (19.92)	0.003**

*Independent *t* test, **Statistically significant. SD: Standard deviation

more scores than other options.^[25] Hence, using a method, in which the significance of the two variables would be determined and assigned the same importance to the two subscales, the final score obtained for the higher option is more realistic and will be lower than the pair-wise choice.^[25] Considering the problems and limitations of the NASA-TLX questionnaire (numerical visual rating scale and no fuzzy and two by two selection), as can be seen, the present study changed the calculations and also the way to choose subscales in comparison with the conventional version, different results were obtained, representing that the new tool is more powerful and more realistic. This is also consistent with the study of Amady *et al.* who used the fuzzy logic in the NASA-TLX questionnaire.^[16]

As mentioned, the task performed in this study was completely mental and had a low physical load, and the task requested in the present study was performed in sitting form; hence, it was expected that this subscale to have the lowest value. The expected results were obtained in both questionnaires. Physical load in ergonomics had the intensity and its absence was meaningless, and the concept of “low physical load” is

always discussed in sitting works, and each sitting task also causes some physical load.^[27,28]

With regard to the main problem of the second part of the NASA-TLX questionnaire, in most cases, participants did not choose the physical demand, and the weight of this subscale in the second part was considered as 0. Ultimately, the mean score of this subscale will be <1 and it significantly differs from the reality and the way of decision-making. In the weighing section of the FBWM-NASA-TLX questionnaire, the participant determines the weight of each subscale by expressing the importance of the weight of each subscale, leading one to identify and apply even the least intensity.

CONCLUSIONS

The present study considered the problems and limitations of the NASA-TLX questionnaire; hence, the FBWM-NASA-TLX questionnaire was designed and evaluated. The results showed that the FBWM-NASA-TLX questionnaire, used for estimating scores and making was more realistic about the workload in the task under study. The NASA-TLX questionnaire was designed to be easy to apply, but, because of aforementioned problems, it is necessary to be redesign.

Limitations

Our work clearly has some limitations. The most important one lies in the fact that lack of examination of this questionnaire in an actual job task, and also lack of evaluation using objective methods such as EEG and ERP: event-related potential due to insufficient financial support. One other limitation of this study was the type of task under consideration, which merely examines the mental load and has a very low physical load that cannot be studied by the combined effect of physical and mental load. Considering the reasons for the study's limitations, it might be suggested that field studies be conducted in the future to better assess the current methodology and compare it with other methods, such as SWOT: Strengths, Weaknesses, Opportunities, and Threats.

Acknowledgment

This study was supported by Shiraz University of Medical Sciences for financial support (grant No. IR.SUMS.REC.1397.942).

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Backs RW, Ryan AM, Wilson GF. Psychophysiological measures of workload during continuous manual performance. *Hum Factors* 1994;36:514-31.
2. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Adv Psychol* 1988;52:139-83.
3. Hart SG, editor NASA-task load index (NASA-TLX); 20 years later.

Archive of SID

Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Los Angeles, CA: SAGE Publications; 2006.

4. Delice EK, Can GFJEOÜMvMFD. An Integrated mental workload assessment approach based on Nasa-TLX and SMAA-2: A case study. *Eskişehir Osmangazi Üniversitesi Mühendislik ve Mimarlık Fakültesi Dergisi*. 2018;26:88-99.
5. Rubio S, Díaz E, Martín J, Puente JM. Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Appl Psychol* 2004;53:61-86.
6. Parasuraman R, Hancock PA. *Adaptive Control of Mental Workload. Stress, Workload, and Fatigue*; 2001.
7. Flemisch FO, Onken RJC, Technology, Work. Open a Window to the Cognitive Work Process! Pointillist Analysis of Man-Machine Interaction. 2002;4:160-70.
8. Loft S, Sanderson P, Neal A, Mooij M. Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. *Human Factors* 2007;49:376-99.
9. Vidulich MA, Tsang PS. Methodological and theoretical concerns in multitask performance: A critique of boles, bursk, phillips, and perdelwitz. *Hum Factors* 2007;49:46-9.
10. Wickens CD. Multiple resources and mental workload. *Hum Factors* 2008;50:449-55.
11. Young MS, Brookhuis KA, Wickens CD, Hancock PA. State of science: Mental workload in ergonomics. *Ergonomics* 2015;58:1-17.
12. Yurko YY, Scerbo MW, Prabhu AS, Acker CE, Stefanidis D. Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. *Simul Healthc* 2010;5:267-71.
13. Zheng B, Jiang X, Tien G, Meneghetti A, Panton ON, Atkins MS. Workload assessment of surgeons: Correlation between NASA TLX and blinks. *Surg Endosc* 2012;26:2746-50.
14. Wiebe EN, Roberts E, Behrend TS. An examination of two mental workload measurement approaches to understanding multimedia learning. *Comput Hum Behav* 2010;26:474-81.
15. Malekpour F, Malekpour A, Mohammadian Y, Mohammadpour Y, Shakarami A, Sheikh Ahmadi A. Assessment of mental workload in nursing by using NASA-TLX. *J Urmia Nurs Midwif Facul* 2014;11:892-9.
16. Amady MM, Raufaste E, Prade H, Meyer JP, Fuzzy-TL. Using fuzzy integrals for evaluating human mental workload with NASA-Task Load indeX in laboratory and field studies. *Ergonomics* 2013;56:752-63.
17. Bridger RS, Brasher K. Cognitive task demands, self-control demands and the mental well-being of office workers. *Ergonomics* 2011;54:830-9.
18. Nygren TE. Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors* 1991;33:17-33.
19. Hart SG. NASA Task load Index (TLX). Volume 1.0; Paper and pencil package. National Aeronautics and Space Administration 1986;2:10-5.
20. Ruiz-Rabelo JF, Rodriguez NE, Di-Stasi LL, Jimenez ND, Bermon CJ, Iglesias DC, *et al.* Validation of the NASA-TLX score in ongoing assessment of mental workload during a laparoscopic learning curve in bariatric surgery. *Obes Surg* 2015;25:2451-6.
21. Kane MJ, Conway AR, Miura TK, Colflesh GJ. Working memory, attention control, and the N-back task: A question of construct validity. *J Exp Psychol Learn Mem Cogn* 2007;33:615-22.
22. Jaeggi SM, Buschkuhl M, Perrig WJ, Meier B. The concurrent validity of the N-back task as a working memory measure. *Memory* 2010;18:394-412.
23. Li RJ. Fuzzy method in group decision making. *Comput Math Appl* 1999;38:91-101.
24. Guo S, Zhao H. Fuzzy best-worst multi-criteria decision-making method and its applications. *Knowl-Based Syst* 2017;121:23-31.
25. Rezaei J. Best-worst multi-criteria decision-making method. *Omega* 2015;53:49-57.
26. Verhaeghen P, Basak C. Ageing and switching of the focus of attention in working memory: Results from a modified N-Back task. *Q J Exp Psychol* 2005;58:134-54.
27. Straker L, Mathiassen SE. Increased physical work loads in modern work—a necessity for better health and performance? *Ergonomics* 2009;52:1215-25.
28. Choi B, Schnall PL, Yang H, Dobson M, Landsbergis P, Israel L, *et al.* Sedentary work, low physical job demand, and obesity in US workers. *Am J Ind Med* 2010;53:1088-101.