Neuroergonomics and Cognitive Engineering, Vol. 167, 2025, 121–132
https://doi.org/10.54941/ahfe1006133

A Subjective and Objective Synchronization Assessment Method of Cognitive Load for the Lunar Exploration

Yang Weiquan[1,2], Li Zengrui[1,2], Tian Ye[3], Liu Yalei[4], Ren Xinyi[1,2], Wang Tianle[1,2], Chen Shanguang[5,6], and Xia Qianchen[1,5]

[1]The Future Laboratory, Tsinghua University, Beijing, 100084, China
[2]School of Design and Art, Beijing Institute of Technology, Beijing, 100081, China
[3]Department of Industrial Engineering, Tsinghua University, Beijing, 100084, China
[4]Institute of Medical Engineering and Translational Medicine, Tianjin University, Tianjin, 300072, China
[5]National Key Laboratory of Human Factors Engineering, Beijing, 100094, China
[6]China Astronaut Research and Training Center, Beijing, 100094, China

ABSTRACT

The cognitive load of astronauts has a large impact on the efficiency of human-machine co-operation, and the reasonable adjustment of astronauts' cognitive load is crucial to guarantee the success of the mission. However, the current cognitive load prediction and assessment methods have the problems of detaching the evoked task (N-back) from the main task and the high latency of subjective load assessment, which affect the accuracy of the prediction and assessment of cognitive load. Therefore, the study proposes a subjective-objective synchronised cognitive load experimental assessment method for typical tasks on the lunar surface, where a dynamic N-back experimental paradigm is designed according to the operation process to induce different levels of cognitive loads for the astronauts, and at the same time, the astronauts are required to complete corresponding NASA-TLX scale pop-ups for real-time subjective load assessment in different operation processes. In the objective assessment, the multimodal signals of the astronauts were collected based on GSR, ECG and PPG for feature extraction. Finally, this study constructs a comprehensive assessment model of human-machine collaborative effectiveness for lunar surface operations based on behavioural performance and cognitive load state and verifies its validity through typical experimental tasks. The experimental assessment method can comprehensively consider the human-machine cooperative ability of astronauts under the influence of multiple factors in the lunar surface special-cause environment, construct the N-back experimental paradigm of the evoked task and the typical task organically combined experiments, and at the same time reduce the latency of the subjective evaluation, realise the synchronous assessment based on subjective and objective physiological and task data, and effectively enhance the accuracy of the prediction and assessment of the cognitive load.

Keywords: Lunar exploration, Cognitive load assessment, Human-machine interaction for complex systems, Human-machine collaboration

INTRODUCTION

The strong radiation (Benaroya, 2017), microgravity (Schuerger *et al.*, 2019; Zhou *et al.*, 2019) and unstructured special environment (Jablonski and Man, 2021; Liu and Huang, 2024) on the lunar surface will have an impact on the physiology and psychology of astronauts, and the characteristics of their human capabilities (Blakely *et al.*, 2016), such as operating force, cognition and perception, will change, and their human-machine manoeuvring capabilities will be constrained.

During lunar exploration missions, astronauts must process large amounts of information and make rapid decisions (Rai *et al.*, 2011), which inevitably puts a strain on their cognitive systems. Cognitive load theory states that an individual's ability to process information is limited (Young *et al.*, 2014). In lunar operations, if the cognitive load of astronauts is too low, it may lead to insufficient perception of the environment and situational awareness, thus affecting the execution of the mission; while too high a cognitive load may trigger physiological and psychological fatigue, reducing operational efficiency and safety. Therefore, rationally adjusting the cognitive load of astronauts is essential to ensure mission success.

Considering the coupling of multiple factors and the superposition of multiple effects in the lunar operating environment, we must adopt a systematic approach to assess and respond to these complexities, and the continuous optimization of the human-computer collaborative operating model ensures that the astronauts are able to work in an environment that is both efficient and safe (Jorgensen, 2010; Pothier *et al.*, 2019). With the continuous development of space exploration technology, this human-centered design thinking will provide more solid theoretical and practical support for lunar surface exploration and drive the cause of space exploration forward (Laraway *et al.*, 2024).

Evaluation Modelling of Human-Machine Collaboration Efficacy

Reasonable human-machine (lunar rover, robotic arm, etc.) collaborative design can maximize the human role and improve task performance (Mueller, 2006; Xue *et al.*, 2023). For the study of human-machine synergy effectiveness model (Kokotinis *et al.*, 2023; Parasuraman and Riley, 1997), it is often constructed from multiple indicators such as task performance, behavioural performance, and human state (Fu *et al.*, 2023). The effectiveness is a combination of task performance and process influences, which not only includes the assessment of astronauts' task performance but also covers the assessment of human state and behavioural performance, and a reasonable human-machine collaborative system design can effectively ensure the safety of astronauts and the successful completion of the lunar surface exploration mission.

Therefore, in this study, various data of the subject astronauts were counted to construct an efficacy model of human-machine collaboration, and the indexes of the model mainly include behavioural performance and human cognitive load indexes. In the behavioural performance indexes, the astronauts' behavioural performance and responsiveness in tasks were assessed by the reaction time and correctness rate of completing the N-back task. In the cognitive load indexes, the astronauts' physiological and psychological states were evaluated to ensure that they are able to perform the mission in an optimal state. Objective data were assessed by fusing multimodal physiological signals for cognitive load assessment, and subjective data were calculated by embedded NASA-TLX scale scores. After collecting the data, the accuracy and stability of the model were evaluated to verify the influence of each factor on the cognitive load of astronauts and explore the mechanism of its influence. At the same time, we compare the dynamic data of astronauts' indicators under the influence of different task difficulties, analyse the correlation of changes among the indicators, and test the reasonableness of the model construction.

Cognitive Load Assessment

Cognitive load (Paas *et al.*, 2003; Schnotz and Kürschner, 2007) is the proportion of information processing capacity or cognitive resources that meets actual needs, where cognitive resources mainly refer to attentional resources and working memory capacity involved in cognitive processes.

In cognitive load assessment experiments, different levels of cognitive load are usually induced and measured by the N-back paradigm (Ni and Ma, 2024; Owen *et al.*, 2005), and the common measurement methods include subjective scale assessment, physiological signal measurement and task performance assessment, and they are usually assessed in a combination of subjective and objective methods. Objective cognitive load is mainly measured by physiological signals, including Electrocardiogram (ECG), Photoplethysmogram (PPG), Galvanic Skin Response (GSR), which has the advantages of sensitivity and objectivity. ECG is a gold-standard method for measuring heart rate, which is influenced by cognitive load. Increased cognitive load typically results in elevated heart rates (Alshanskaia *et al.*, 2024). Studies have shown that heart rate increases with task difficulty (He *et al.*, 2022; Hettiarachchi *et al.*, 2018), indicating higher cognitive load. For instance, during a cognitive task, heart rate measured via ECG was found to correlate with anxiety and depression levels. PPG measures blood volume changes and can track cognitive load through pulse wave amplitude (Gupta *et al.*, 2024; Xuan *et al.*, 2020). PPG morphometrics, such as changes in waveform shape, are indicative of cognitive load and can be used to differentiate between low and high mental workload (Pavlov *et al.*, 2023). GSR measures skin conductance, which increases with cognitive load due to heightened sympathetic nervous system activity (Feradov *et al.*, 2020; Saha *et al.*, 2022). GSR data has been shown to correlate with task difficulty and cognitive load (Hirachan *et al.*, 2022), providing a reliable measure of mental effort.

Measurement of subjective load is commonly performed by means of a scale, the NASA-TLX Scale (Hart and Staveland, 1988) is a subjective load assessment tool developed by NASA Ames Research Center that measures operator load while performing a task through six dimensions (psychological demands, physical demands, time demands, performance, effort, and frustration). The scale uses a weighted average to produce an overall workload score and is widely used in aerospace, command and control, and other fields.

Experimental Paradigm Improvement Design

The current experimental design of cognitive load assessment has the following major problems. Firstly, the current N-back experimental paradigm (Pergher *et al.*, 2018) often induces different levels of cognitive load through independent visual or auditory stimuli, which is often detached from the content of the main task that the subject is currently performing, affecting the subject's immersive experience in the simulated environment. Secondly, the subjective load scale is often used some time after the end of the experiment to make the subject recall the experimental process results to answer, resulting in high assessment latency, which seriously affects the accuracy of cognitive load assessment.

Therefore, the study proposes a subjective and objective synchronized cognitive load experimental assessment method for the lunar soil acquisition mission, where a dynamic N-back experiment paradigm is designed based on the operation process to induce different levels of cognitive loads for astronauts, and at the same time, astronauts are required to complete the corresponding modular NASA-TLX scale pop-ups for real-time subjective load assessment in different operation processes. In the objective assessment, the GSR, ECG and PPG of the astronauts were collected for feature extraction based on the multi-probe physiological module.

To collect physiological data from subjects under different cognitive loads, a within-group experimental design was used, with the within-group factor being the difficulty of the task, which was categorized into three levels of difficulty: low, medium, and high. The experiment was conducted as a simultaneous primary and secondary task, with the primary task being moon soil collection, which required the shovel to be manipulated by keystrokes to collect 150g of moon soil six times. The collected components were set to be volcanic glass beads, the average diameter is generally less than 0.1mm, which is important for the study of early lunar magma, and the colors of volcanic glass beads of different compositions are significantly different, as shown in the Table 1.

Table 1: Volcanic glass moon soil color classification and characteristics.

Color	Feature
Green	TiO_2 content is the lowest (generally less than 1%),
Orange	TiO_2 content up to 9%~12%
Yellow	Slightly higher TiO_2 content (3%~7%)
Black	Highest TiO_2 content, above 14%

Based on the specificity of the color of moon soil collection, the subtask was designed as an improved N-back task, which required subjects to remember the number of times and the color of the current collection after each collection of moon soil samples, and to pop up a pop-up window during the subsequent collection asking whether the color of the current collection was the same as the color of the specific number of times, in order to determine the behavioral performance of the subjects during the experiment.

Study Goal

The research was based on the virtual reality simulation of the lunar surface training environment and synthesizes multi-physiological signals to realize the subjective and objective simultaneous assessment of astronauts' cognitive load. This method can comprehensively consider the human-machine synergistic ability of astronauts under the influence of multi-factors in the lunar surface special cause environment, construct the N-back experimental paradigm of the evoked task with the typical task organically combined with the experiments, and at the same time reduce the latency of the subjective evaluation, realize the synchronous evaluation based on the subjective and objective physiological and task data, which effectively improves the accuracy of the prediction and evaluation of the cognitive load, and lastly, combined with the study of the astronauts' human-machine synergistic efficacy in lunar exploration, we can evaluate the impact of cognitive load in the mission. Finally, we analyse the elements affecting the human-machine synergistic performance in the light of the research on the human-machine synergistic performance of astronaut lunar surface exploration.

METHOD

Participants

In order to validate the usability of this experimental system while collecting physiological datasets under different loads, a total of 10 participants (Virzi, 1992) were recruited for the experiment. The average age among the participants was 22.5 ($SD = 0.92$). The experiments were conducted at the Beijing Institute of Technology and were reviewed by the Ethical Review Board of BIT.

Material and Procedure

Based on a self-developed virtual reality (VR) lunar surface environment, this study simulates the operational task of lunar soil collection, and employs physiological multi-parameter modules (oxygen clamps, electrocardiographic electrodes, and electrocorticographic electrodes) to assess the cognitive load of astronauts. The experiment induced different levels of cognitive load in astronauts through a dynamic N-back task, incorporating a pop-up real-time NASA-TLX scale.

The dynamic graphic N-back task is designed as follows: the subject needs to collect moon soil six times in total, and after shovelling out the moon soil, the colour of some of the volcanic glass bead samples can be seen. 1-back task, the subject needs to press the handle button as fast as possible when he/she sees that the colour of the shovelled moon soil is the same as that of the last collection. 2-back task, the subject needs to press the handle button as fast as possible when he/she sees that the colour of the shovelled moon soil is the same as that of the second collection (two times in between). In the 2-back task, subjects had to press the handle button as fast as they could when they saw that the lunar soil they were shovelling was the same colour as that of the second previous collection (one interval). 3-back task, subjects had to press the handle button as fast as they could when they saw that the lunar soil they were shovelling was the same colour as that of the third previous collection (two intervals). The software platform recorded the time from when the digging sample saw the colour to when the key was pressed as the response time each time. The result of each response was also recorded and used to count the percentage of correctness. If the subject did not respond for more than five seconds, it was recorded as one failure.

After the n-back question in times 4–6, we included a NASA-TLX scale pop-up to embed the scale during the task to enhance timeliness, as shown below.

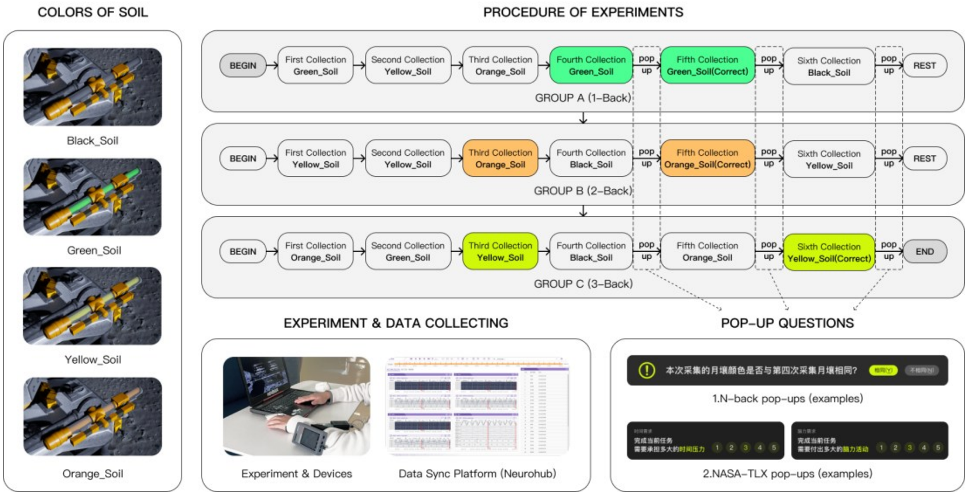

Figure 1: Procedure and environment of the experiment.

Data Processing and Analysis

For objective physiological data, we used the open-source Neurokit2 python library to preprocess and feature extract the signals during the N-back task. The sampling rates for ECG, GSR, and PPG were set at 1000 Hz, 100 Hz, and 80 Hz, to ensure high-resolution capture of physiological signals for detailed analysis. To capture the dynamic changes in astronauts' cognitive load in real time, a sliding time window of 30 seconds was employed to segment the data (Xia *et al.*, 2020).

The original data is first filtered to reduce noise. A 0.5 Hz high-pass Butterworth filter is applied to the ECG signal, followed by the removal of power line interference. A band-pass Butterworth filter with cutoff frequencies of 0.5 Hz and 3 Hz is used to filter the respiration signal. We extracted heart rate (HR) by identifying R-peaks in the ECG signal and calculating the number of R-peaks per minute. For heart rate variability (HRV) analysis, we computed the standard deviation of the RR intervals (SDNN) and the root mean square of successive differences (RMSSD) to quantify the variability in heart rate (Laraway *et al.*, 2024).

For the GSR data, a second-order Butterworth filter with a cutoff frequency of 0.3 Hz was employed to effectively remove high-frequency noise and smooth the signal. Given that the useful frequency band of GSR signals primarily lies below 0.2 Hz, this filter design ensures maximum flatness in the passband and a gradual attenuation in the stopband, thereby effectively eliminating out-of-band noise. For segments of the signal that are entirely submerged in noise, these valueless data points were discarded to ensure the quality and reliability of the processed GSR data. Skin Conductance Level (SCL) was extracted by calculating the average value of the tonic component of the filtered GSR signal within each time window, reflecting the baseline level of skin conductance. The amplitude and the height of Skin Conductance Responses (SCR) was determined by identifying the peaks in the phasic component of the filtered GSR signal, with the SCR representing the magnitude of the phasic responses.

During the processing of PPG data, a third-order Butterworth filter with a passband frequency range of 0.5 to 8 Hz was utilized to eliminate noise and interference (Alshanskaia *et al.*, 2024). The purification and peak detection of PPG signals were accomplished by leveraging the integrated functions of NeuroKit2 and the moving average method was employed to identify potential systolic peaks. We measured the PPG amplitude by identifying the peak values of the pulse waves, which reflect the intensity of blood flow and vascular tone.

In the calculation of the scores of the NASA-TLX scale, the weight values of the dimensions were obtained by two-by-two comparisons, and the product of everyone's score and weight was used as the subjective cognitive load score.

RESULTS

The distribution of physiological features across different n-back task levels is illustrated in Figure 2. A one-way analysis of variance (ANOVA) revealed that SCL increased significantly with higher task difficulty ($p < 0.01$). Specifically, both 3-Back and 2-Back exhibited significantly higher SCL values compared to 1-Back ($p < 0.01$), though no significant difference was observed between 3-Back and 2-Back.Conversely, PPG Amplitude and HRV decreased progressively with increasing task demands. PPG Amplitude declined from 1-Back to 3-Back, showing a significant negative correlation with task difficulty ($p < 0.01$). Similarly, HRV decreased from 1-Back to 3-Back ($p < 0.01$) and 2-Back to 3-Back ($p < 0.05$), indicating reduced parasympathetic modulation under high cognitive load. These results suggest that this experimental design is effective in inducing three levels of cognitive load, high, medium and low, and that an increase in cognitive load induces an increase in sympathetic neural activity and a decrease in cardiovascular adaptations

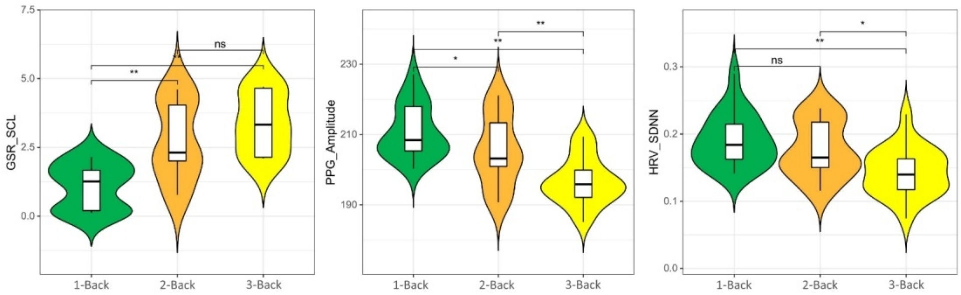

Figure 2: Feature extraction results.

In the assessment of subjective workload and behavioural performance, the NASA-TLX value and accuracy rate of the N-back for all participants across different groups and response times are illustrated in Figure 3. A one-way ANOVA revealed that the subjective workload increased significantly with the increasing number of n in the n-back task ($p < 0.01$). The average scores across the three groups were as follows: 3-Back ($M = 14.41$, $SD = 0.72$) > 2-Back ($M = 12.88$, $SD = 0.61$) > 1-Back ($M = 8.52$, $SD = 0.43$). Additionally, the subjective workload scores of participants increased significantly over time as the task progressed ($p < 0.01$).

Using the accuracy rate of the N-back as a reference for the behavioural performance of participants in the lunar surface simulation task, it was observed that the task performance of participants decreased over time, with the most significant decline observed in the 2-back and 3-back groups. The average accuracy rates across the three groups were: 1-back (86.67%) > 2-back (66.67%) > 3-Back (50%).

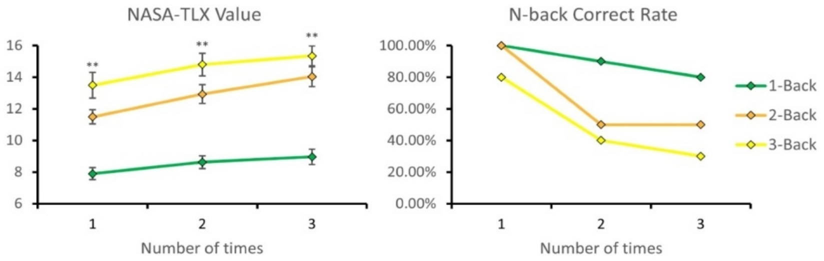

Figure 3: NASA-TLX value and N-back correct rate.

DISCUSSION

The findings of this study demonstrate a clear decline in behavioural performance, as measured by the accuracy of the N-back task, with increasing cognitive load. This trend is consistent across all three experimental groups. Specifically, as task difficulty increased from 1-back to 3-back levels, a significant reduction in accuracy was observed. This decline in performance corroborates with subjective cognitive load data derived from NASA-TLX scores and objective physiological measurements obtained from ECG, GSR, and PPG signals. These results reinforce the hypothesis that higher task demands adversely affect working memory and attentional resources, leading to degraded task performance.

A noteworthy contribution of this study is the robust synchronization observed between subjective cognitive load assessments and objective physiological data. The real-time NASA-TLX pop-up scales provided a timely capture of subjective load perceptions, effectively minimizing latency-related inaccuracies typically associated with retrospective reporting. Simultaneously, the multimodal physiological signals offered granular insights into the astronauts' cognitive states, with features such as heart rate variability (HRV) from ECG, skin conductance levels (SCL) from GSR, and pulse amplitude variability from PPG showing high sensitivity to task difficulty. The alignment of these subjective and objective metrics underscores the reliability and validity of the proposed synchronized assessment framework.

An important observation is the temporal degradation of behavioural performance over the duration of the tasks. This decline was accompanied by a progressive increase in both subjective and objective cognitive load indicators. The temporal analysis revealed that participants exhibited reduced accuracy and prolonged reaction times in later stages of the task sequence, reflecting cumulative cognitive fatigue. Physiological measures such as elevated heart rates and increased skin conductance levels further validated the temporal escalation in cognitive load. These findings highlight the dynamic nature of cognitive load, emphasizing the need for adaptive workload management strategies during prolonged operations in high-stakes environments like lunar exploration.

CONCLUSION AND FUTURE WORK

This study verifies the validity of the subjective-objective synchronization assessment method and the human-computer cooperative effectiveness model through usability experiments. This human-computer cooperative technology experimental method can be widely applied to human-computer interaction environments in complex systems such as human spaceflight and autonomous driving. In further research, the number of subjects will be increased to obtain more reliable data sets, and the cognitive load will be predicted by machine learning classification algorithms. At the same time, the model parameters will be adjusted to improve the accuracy of the system, and the generalization ability of the subjective and objective assessment model of cognitive load will be improved so that it can be adapted to different individual astronauts and different deep space exploration mission requirements. In addition, exploring more covert physiological measurement techniques and reducing the impact of acquisition equipment on astronaut operations are also our further research goals.

ACKNOWLEDGMENT

This work was supported by the Beijing Natural Science Foundation of China (Grant No. 3254040), and National Key Laboratory of Human Factors Engineering.

REFERENCES

Alshanskaia, E. I., Zhozhikashvili, N. A., Polikanova, I. S. and Martynova, O. V. (2024), "Heart rate response to cognitive load as a marker of depression and increased anxiety", *Frontiers in Psychiatry*, Vol. 15, p. 1355846, doi: 10.3389/fpsyt.2024.1355846.

Benaroya, H. (2017), "Lunar habitats: A brief overview of issues and concepts", *REACH*, Vol. 7–8, pp. 14–33, doi: 10.1016/j.reach.2018.08.002.

Blakely, M. J., Wilson, K., Russell, P. N. and Helton, W. S. (2016), "The impact of cognitive load on volitional running", presented at the Proceedings of the Human Factors and Ergonomics Society, pp. 1178–1182, doi: 10.1177/1541931213601276.

Feradov, F., Ganchev, T. and Markova, V. (2020), "Automated Detection of Cognitive Load from Peripheral Physiological Signals based on Hjorth's Parameters", *2020 International Conference on Biomedical Innovations and Applications (BIA)*, presented at the Proceedings, 2020, pp. 85–88, doi: 10.1109/BIA50171.2020.9244287.

Fu, Y., Wang, H., Xue, S., Zhang, B., Cheng, J. and Gao, Z. (2023), "A new paradigm of human-robot dialogue facilitating human and robot teaming in sample exploration", presented at the Proceedings of the 2023 IEEE International Conference on Real-Time Computing and Robotics, RCAR 2023, pp. 805–810, doi: 10.1109/RCAR58764.2023.10249356.

Gupta, S., Gupta, K. and Singh, A. (2024), "Automated assessment of mental workload from PPG sensor data using cross-wavelet coherence and transfer learning", *Biomedical Engineering Letters*, Vol. 14 No. 4, pp. 891–902, doi: 10.1007/s13534-024-00384-1.

Hart, S. G. and Staveland, L. E. (1988), "Development of NASA-TLX (task load index): Results of empirical and theoretical research", in Hancock, P. A. and Meshkati, N. (Eds.), *Advances in Psychology*, Vol. 52, North-Holland, pp. 139–183, doi: 10.1016/S0166-4115(08)62386-9.

He, D., Wang, Z., Khalil, E. B., Donmez, B., Qiao, G. and Kumar, S. (2022), "Classification of Driver Cognitive Load: Exploring the Benefits of Fusing Eye-Tracking and Physiological Measures", *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2676 No. 10, pp. 670–681, doi: 10.1177/03611981221090937.

Hettiarachchi, I., Hanoun, S., Nahavandi, D., Iskander, J., Hossny, M. and Nahavandi, S. (2018), "Towards More Accessible Physiological Data for Assessment of Cognitive Load - A Validation Study", *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, presented at the Proceedings - 2018, pp. 3045–3050, doi: 10.1109/SMC.2018.00517.

Hirachan, N., Mathews, A., Romero, J. and Rojas, R. F. (2022), "Measuring Cognitive Workload Using Multimodal Sensors", *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBS)*, Vol. 2022-July, presented at the Proceedings, pp. 4921–4924, doi: 10.1109/EMBC48229.2022.9871308.

Jablonski, A. M. and Man, K. F. (2021), "Impact of recent lunar missions on the understanding of lunar environment", presented at the Earth and Space 2021: Space Exploration, Utilization, Engineering, and Construction in Extreme Environments - Selected Papers from the 17th Biennial International Conference on Engineering, Science, Construction, and Operations in Challenging Environments, pp. 931–943, doi: 10.1061/9780784483374.085.

Jorgensen, J. (2010), "Humans: The strongest and the weakest joint in the chain", *Lunar Settlements*, pp. 247–259, doi: 10.1201/9781420083330.

Kokotinis, G., Michalos, G., Arkouli, Z. and Makris, S. (2023), "On the quantification of human-robot collaboration quality", *International Journal of Computer Integrated Manufacturing*, Vol. 36 No. 10, pp. 1431–1448, doi: 10.1080/0951192X.2023.2189304.

Laraway, S., Snycerski, S., Pradhan, S., Huitema, B., Rantz, W., Whitehurst, G. and Battiste, V. (2024), "An Introduction to Single-Case Experimental Designs for Applied Human Factors and Ergonomics", *Neuroergonomics and Cognitive Engineering*, Vol. 126, presented at the AHFE (2024) International Conference, AHFE Open Acces, doi: 10.54941/ahfe1004742.

Liu, Y. and Huang, S. (2024), "In situ measurements of thermal environment on the moon's surface revealed by the chang'E-4 and chang'E-5 missions", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 17, pp. 2149–2160, doi: 10.1109/JSTARS.2023.3340853.

Mueller, R. P. (2006), "Surface support systems for co-operative and integrated human/robotic lunar exploration", Vol. 3, presented at the AIAA 57th International Astronautical Congress, IAC 2006, pp. 1871–1879.

Ni, L. and Ma, W. J. (2024), "A computational approach to the N-back task", *Scientific Reports*, Vol. 14 No. 1, p. 30211, doi: 10.1038/s41598-024-80537-5.

Owen, A. M., McMillan, K. M., Laird, A. R. and Bullmore, E. (2005), "N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies", *Human Brain Mapping*, Vol. 25 No. 1, pp. 46–59, doi: 10.1002/hbm.20131.

Paas, F., Renkl, A. and Sweller, J. (2003), "Cognitive load theory and instructional design: Recent developments", *Educational Psychologist*, Vol. 38 No. 1, pp. 1–4, doi: 10.1207/S15326985EP3801_1.

Parasuraman, R. and Riley, V. (1997), "Humans and automation: Use, misuse, disuse, abuse", *Human Factors*, Vol. 39 No. 2, pp. 230–253, doi: 10.1518/001872097778543886.

Pavlov, Y. G., Gashkova, A. S., Kasanov, D., Kosachenko, A. I., Kotyusov, A. I. and Kotchoubey, B. (2023), "Task-evoked pulse wave amplitude tracks cognitive load", *Scientific Reports*, Vol. 13 No. 1, p. 22592, doi: 10.1038/s41598-023-48917-5.

Pergher, V., Wittevrongel, B., Tournoy, J., Schoenmakers, B. and Van Hulle, M. M. (2018), "*N*-back training and transfer effects revealed by behavioral responses and EEG", *Brain and Behavior*, Vol. 8 No. 11, p. e01136, doi: 10.1002/brb3.1136.

Pothier, B., Schlacht, I. L., Musilova, M., Alcibiade, A., Foing, B. and Rogers, H. (2019), "A case study of human factor & anthropological investigations in space mission simulations and analogs", Vol. 2019-October, presented at the Proceedings of the International Astronautical Congress, IAC.

Rai, B., Kaur, J., Foing, B. H. and O'Griofa, M. (2011), "JBR group study of biomedical experiments results: MDRS 100B ILEWG EuroMoonMars crew", Vol. 1, presented at the 62nd International Astronautical Congress 2011, IAC 2011, pp. 168–172.

Saha, S., Jindal, K., Shakti, D., Tewary, S. and Sardana, V. (2022), "Chirplet transform-based machine-learning approach towards classification of cognitive state change using Galvanic skin response and photoplethysmography signals", *Expert Systems*, Vol. 39 No. 6, p. e12958, doi: 10.1111/exsy.12958.

Schnotz, W. and Kürschner, C. (2007), "A reconsideration of cognitive load theory", *Educational Psychology Review*, Vol. 19 No. 4, pp. 469–508, doi: 10.1007/s10648-007-9053-4.

Schuerger, A. C., Moores, J. E., Smith, D. J. and Reitz, G. (2019), "A lunar microbial survival model for predicting the forward contamination of the moon", *Astrobiology*, Vol. 19 No. 6, pp. 730–756, doi: 10.1089/ast.2018.1952.

Virzi, R. A. (1992), "Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough?", *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 34 No. 4, pp. 457–468, doi: 10.1177/001872089203400407.

Xia, Q., Lv, J., Ma, S., Gao, B. and Wang, Z. (2020), "A new information-theoretic method for advertisement conversion rate prediction for large-scale sparse data based on deep learning", *Entropy*, Vol. 22 No. 6, p. 643, doi: 10.3390/e22060643.

Xuan, Q., Wu, J., Shen, J., Ji, X., Lyu, Y. and Zhang, Y. (2020), "Assessing cognitive load in adolescent and adult students using photoplethysmogram morphometrics", *Cognitive Neurodynamics*, Vol. 14 No. 5, pp. 709–721, doi: 10.1007/s11571-020-09617-2.

Xue, S., Liao, G., Tan, L., Tian, Y., Wu, Y., Fu, Y., Zhang, Z., *et al.* (2023), "Research on human-robot cooperative target recognition for spatial sampling task", Vol. 941 LNEE, presented at the Lecture Notes in Electrical Engineering, pp. 434–441, doi: 10.1007/978-981-19-4786-5_60.

Young, J. Q., Van Merrienboer, J., Durning, S. and Ten Cate, O. (2014), "Cognitive load theory: Implications for medical education: AMEE guide no. 86", *Medical Teacher*, Vol. 36 No. 5, pp. 371–384, doi: 10.3109/0142159X.2014.889290.

Zhou, G., Li, R., Yan, K., Zhao, X., Chen, J., Mo, P. and Wang, G. (2019), "Theoretical and experimental methods for lunar regolith/rock related mechanical issues in lunar minerals mining", *Meitan Xuebao/Journal of the China Coal Society*, Vol. 44 No. 1, pp. 183–191, doi: 10.13225/j.cnki.jccs.2019.0029.