Research Article

Quantifying Minute-by-Minute Modeling of Cognitive Fatigue in a Stroop Task: Slower Exponential Decay and the Effect of a Brief Rest

Jeong-Pyo Han

Saint James School, 17641 College Rd, Hagerstown, MD 21740, United States

ABSTRACT

Sustained attention usually deteriorates during hour-scale tasks. However, it remains unclear whether this attention decline is the same at a minute-by-minute scale, and also if there would be a benefit of having very short rests. This study specifically seeks to answer which simple fatigue process best identifies performance over a 60-minute Stroop-type task, and whether having a brief mid-task rest significantly improves the performance of a task. This study hypothesized that a slower exponential decline of attention would fit better than a standard or faster decline, and also that adding a one-minute passive rest with a small post-break boost would improve task performance without affecting the fit before the break. This study implemented four fatigue regimes (standard, slower, faster, and recovery) as one-minute-step exponential models with Gaussian observation noise. Grid search was conducted to set parameters by minimizing mean squared error against the benchmark, while estimating uncertainty for error metrics with bootstrap confidence intervals and the one for correlation with Fisher-z intervals. For each regime, this study ran R=10,000 simulations (seed=2025). Run-mean trajectories were reported with 95% pointwise bands, while conducting sensitivity analyses varying the fatigue rate, noise level, recovery magnitude, and break timing. Results in this study supported the hypothesis that the slower model provided the best overall performance (e.g. r=0.89, mean squared error = 0.0035), followed by the recovery model (r=0.86, mean squared error =0.0043), and the faster model (r=0.71, mean squared error = 0.0117). These findings were supported by the sensitivity analysis that deviating fatigue rate from its fitted value degraded the fit of the model, while higher noise widened bands without increasing means. In addition, more end-of-task performance was preserved by earlier breaks.

Keywords: Cognitive fatigue; sustained attention; Stroop task; exponential decay; time-on-task

INTRODUCTION

Cognitive fatigue is defined as a prevalent and potentially debilitating symptom, impacting the life of individuals diagnosed with an acute and chronic medical condition (1). The etiological mechanism of fatigue is currently not fully understood. Fatigue is a complex, multidimensional phenomenon that occurs independently of specific medical diagnoses. However, its distinctive components can be reliably assessed using validated clinical instruments (2). However, the majority of existing studies conducted on cognitive fatigue dealt with physiological measurements without providing simulation-based applications of the results for a more generalized framework (3). To be more specific, despite the increased literature of cognitive

Corresponding author: Jeong-Pyo Han, E-mail: hanjeongpyo96@gmail.com.

Accepted September 17, 2025
https://doi.org/10.70251/HYJR2348.35400406

fatigue, most of the studies were conducted in a clinical setting with limited observations of patients. Several studies (4), (5) generated a computational model of cognitive fatigue, but they did not incorporate the rest and recovery process even though there was evidence of even short breaks significantly improving the metabolic conditions from fatigue (6). There is currently a lack of replicated mathematical simulations addressing cognitive fatigue as cognitive fatigue is only partially clinically observable (7).

This paper aims to help bridge this literature gap by developing four distinct quantitative models to simulate cognitive fatigue over time, validating existing findings of the studies and re-simulating time-on-task effects using Python. In order to incorporate real-world cognitive task demands, this study applies time-on-task effects that are known to progressively decline when people are engaged in a task over time (8). These effects are particularly important to consider when developing strategies to mitigate fatigue-related risks. Prior research, for example, has examined time-on-task in contexts, including student engagement and long-distance driving (9), (10). However, despite its broad application, the methodologies for simulating time-on-task are often underreported (11). To address these issues, this study incorporates a standard fatigue model, a slower fatigue model, a faster fatigue model, and a recovery fatigue model as four distinct mathematical models under various conditions through a sustained attention task named Stroop Task. Ultimately, this study seeks to explore how cognitive performance is influenced by different fatigue decay rates and recovery mechanisms.

METHODS AND MATERIALS

Research Question and Hypothesis

This study aims to answer the research question and test the hypothesis as follows.

Research Question: Which simple fatigue process among standard exponential decay, a slower or faster decay, or a recovery-integrated model with a brief mid-task break, best matches minute-by-minute accuracy on a 60-minute Stroop benchmark, and does brief rest improve post-break performance without affecting pre-break fit?

Hypothesis: The slower exponential decay model will present the best overall fit to the benchmark on a minute-by-minute Stroop task during 60 minutes compared to the standard and faster models.

Mirroring the experimental research with scalability and flexibility of simulation, the goal was to generate a quantitative model to replicate performance decline from cognitive fatigue. After generating four different models based on fatigue progression: standard fatigue, slower fatigue, faster fatigue, and recovery integrated condition, these models were tested against benchmark data provided (4) which analyzed Stroop task as a reliable means in measuring sustained attention and fatigue effects.

Through a continuous attention-sustained Stroop task, this study simulated task engagement that spanned for 60 minutes, while recording performance values at one-minute intervals. For comparison purposes, the standard fatigue model was simulated with typical exponential decay, the slower fatigue model was applied with more gradual performance decline, the faster fatigue model was applied with more rapid decline, and the recovery-integrated model was applied with a rest period in the middle of the task through temporary restoration in the performance. This multimodal design in this study allowed for direct comparison of cognitive fatigue in different situations, and identified key indicators that affected cognitive fatigue.

Simulation Model and Equation

Performance at minute t was modeled as $y_t = \alpha + \gamma e^{-kt} + \varepsilon_t$, where α is the baseline, γ is the amplitude, k is the fatigue rate, and ε_t is Gaussian noise (mean 0, standard deviation 0.03). This model applied least squares for the benchmark, allowing the starting level and long-run level to be determined directly by the data instead of the fixed in advance.

Parameter Fitting and Core Settings

The k value was selected by minimizing the mean squared error (MSE) (to be around 0.030 in the data). For milder and steeper declines, the k values were used for 0.015 for “Slower” and 0.050 for “Faster.” The noise size used in simulations was the same as the standard deviation of residuals between the benchmark and the best-fit exponential at the chosen value of k (≈ 0.03). (Table 1).

Regime Specifications

For the standard fatigue model, a general fatigue model simulated a general fatigue condition expected in average sustained attention tasks. For a slower fatigue model, a lower decay rate was applied to simulate greater cognitive endurance. For a faster fatigue model,

Table 1. Model Parameters and Run Settings Used in Simulations

Setting	Value (Standard otherwise noted)	How chosen
Fatigue rate k	0.030 (Standard); 0.015 (Slower); 0.050 (Faster)	Grid search minimizing mean squared error; simple offsets for slower/faster
Noise standard deviation	≈ 0.03	Residual SD from deterministic best-fit curve
Recovery boost	0.08	One-minute passive rest effect
Break minute	30	Recovery regime only
Runs per regime	10,000	Stability of bands/intervals
Random Seed	2025	Exact reproducibility

k = fatigue-decay constant (standard 0.030; slower 0.015; faster 0.050). Noise standard deviation was around 0.03 that was used as the residual standard deviation from the best-fit exponential curve to the Stroop benchmark. Δ indicated the one-time recovery boost that was only used in the recovery regime. Mid-task 1 minute rest at minute 30 was the break minute. R = 10,000 runs (seed = 2025) were used for each regime to calculate minute-wise mean trajectories and 2.5^{th} to 97.5^{th} percentiles. Accuracy was capped in [0, 1].

higher decay rate was applied to simulate individuals with rapidly building fatigue from stress-intensive environments. Gaussian noise with mean 0 and standard deviation 0.03 was added to each time step for realistic simulation. For the recovery-integrated model, a piecewise approach was used. For example, when t is less than or equal to 30, the standard exponential decay k-value of 0.03 was used. For t to be between 30 and 31 minutes, there was a recovery increment of 0.08 applied during a rest interval. For t to be greater than 31 minutes, continuous exponential decay was applied to the performance baseline. In the recovery regime, the task was paused at minute 30 for one minute as a passive rest. Immediately after the pause, a one-time recovery boost Δ = 0.08 was applied (e.g., 0.08 was added to the current accuracy, capped at 1.0), and the same decay rate k was used before and after the break.

Software and Reproducibility

With this study design, simulation was implemented with a Python program through the Numpy for calculation of numerical values and Matplotlib libraries for the generation of visuals. This study ran R = 10,000 independent simulations per regime with fixed random seed 2025. Following each step, values were capped to the possible range from 0 to 1. The minute-wise mean trajectory was reported, while indicating the 95% pointwise band (2.5^{th}– 97.5^{th} empirical percentiles across runs). Simulation was performed in a one-minute interval as a discrete step to apply the dynamics of hour-long cognitive engagement, and this allowed the analysis of initial task engagement, mid-task performance, recovery effects, and end-task fatigue conditions.

Sensitivity Analysis

The model was stress-tested by varying one factor at a time, while holding all other conditions at their standard values described above. Then, R=10,000 simulations were rerun per setting (seed 2025). Specifically, fatigue rate (k) was calculated by multiplying {0.5, 0.75, 1.00, 1.25, 1.50} by the fitted Standard k. Noise (standard deviation) was calculated to be {0.015, 0.030, 0.045} (i.e. {0.5, 1.0, 1.5}* 0.03). Recovery size (Δ) was {0.04, 0.08, 0.12}. Break timing was {20, 30, 40} minutes (recovery regime only). For each setting, r, mean squared error, and mean absolute error were calculated, along with their 95% confidence intervals using the same procedures: bootstrap for error metrics; Fisher-z for r). Results were considered to be strong when the change relative to the Standard setting was $|\Delta| \leq 0.03$, and mean squared error changes were calculated to be less than or equal to 10%. Data were recorded in each simulation at one-minute intervals for each simulation, enabling the alignment with benchmark data through residual analysis.

Benchmark Dataset and Evaluation Metrics

Benchmark data were used from the results in the study (4) which analyzed Stroop task as a reliable means in measuring sustained attention and fatigue effects. With this benchmark data, simulation scale was matched with averaged and normalized performance accuracy and reaction times to produce a benchmark curve for comparison with fatigue progress-based models in this study. Output from each model was overlaid with the benchmark Stroop curve for visual comparison. In addition, mean squared error provided a numeric measure of model fit, while similarity was quantified with Pearson correlation coefficient. Time-specific discrepancies over the full 60-minute task engagement were indicated by residual plots. For each regime, Pearson correlation r was reported along with the benchmark, mean squared error, and mean absolute error. Nonparametric bootstrap was used over run (10,000 resamples) to estimate the uncertainty of mean squared error/mean absolute error. Then, for r, the standard Fisher-z method was used to compute 95% confidence intervals.

RESULTS

Through the simulation, the standard fatigue model provided a downward trend across the 60-minute interval, and the performance declined gradually from 1.0 to about 0.15 by the minute 60. For the first 20 minutes, the model showed a similar pattern with Stroop benchmark, but after 30 minutes, increasing residuals were recorded, indicating that the fatigue effects were being overestimated by the model. Correlation coefficient with the benchmark data was r=0.82, and mean squared error was 0.0062.

For the faster fatigue model, a sharp decline was shown with gradually decreasing performance below 0.5 by 25 minutes. At the end of the 60-minutes task, it reached almost zero. Performance levels were underestimated significantly by this model as observed in Stroop data. This suggested that cognitive fatigue progression was not accurately reflected by a high decay constant. The correlation coefficient was r=0.71, and the mean squared error was 0.0117.

For the slower fatigue model, gradual performance decline was shown, reaching 0.45 by 60 minutes, and closely followed the Stroop benchmark in the entire interval. The highest correlation coefficient was r=0.89 among all models, but the lowest mean squared error was calculated as 0.0035. This suggested the strongest result derives throughout an hour of task engagement. In addition, minimum deviation was shown in residual plots, not exceeding 0.05.

For the recovery-integrated model, performance increased by 0.08 at around 30 minutes from the simulated rest period, and decay resumed afterwards. This transient improvement aligns with the results in previous study (12) which analyzed task-induced fatigue recovery. Correlation coefficient with the benchmark data was r=0.86, and the mean squared error was 0.0043. Only a small divergence around the break was recorded with the simulated recovery adjustment, demonstrating that even a short break can delay fatigue progression, while sustaining high performance levels in longer periods. The performance decline was shown over 60-minute task engagement (Figure 1), and residuals of standard, faster, and recovery models were compared to the Stroop benchmark (Figure 2). A summary of fit metrics for all four regimes (r, MSE, and end-of-task accuracy) was reported in Table 2.

In summary of findings, all three models showed a similar pattern for the first 20 minutes and aligned well with the benchmark data. Zooming into the first 20 minutes indicated that all four regimes followed the benchmark closely early on (Figure 3). The faster model started diverging within the first quarter hour, while the slower and recovery models were nearly indistinguishable from the benchmark, suggesting that short tasks may mask differences in fatigue rate. In an interval between 20 and 40 minutes, the faster fatigue model started deviating sharply, while the standard fatigue model began to decline. Slower fatigue model

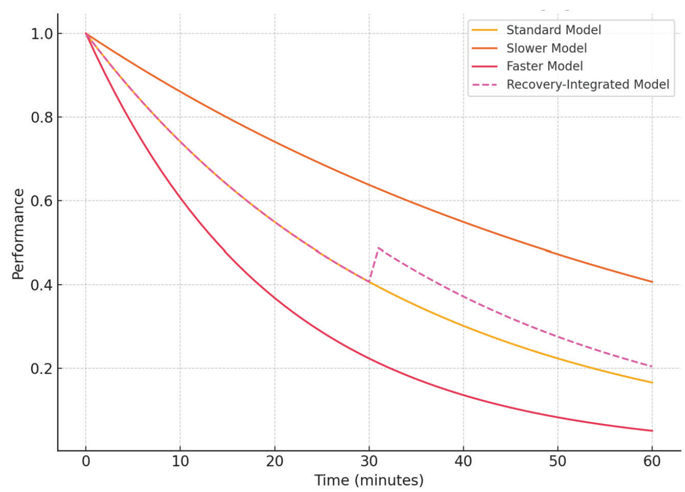

Figure 1. Mean Simulated Accuracy Over a 60-Minute Task for Four Regimes with Uncertainty. Mean accuracy on a minute-by-minute basis was shown for all four Standard, Slower, Faster, and Recovery-integrated models. Shared areas were 95% pointwise bands across R = 10,000 runs. At minute 30, the vertical market was shown, indicating the 1-minute rest (only for recovery regime), followed by a one-time boost $\Delta = 0.08$. Accuracy was capped in [0 1]. The benchmark Stroop curve was indicated for visual comparison.

aligned closely with the benchmark data. Furthermore, after the simulated rest interval, there was a noticeable rebound in the recovery-integrated model. Residuals increased sharply in standard and faster fatigue models in the interval of 40 to 60 minutes, while residuals of the slower fatigue and recovery-integrated models maintained the close alignment with the benchmark curve.

For confidence interval analysis, with repeated simulations in this study, faster fatigue model showed greater variability with wider error range, while the slower fatigue and recovery-integrated model showed narrow confidence intervals, suggesting strong evidence of stability. In addition, long-term predictions were shown to be significantly influenced by small changes in k values or recovery magnitude through sensitivity analysis. Scaling k away from its fitted value degraded fit monotonically, while uncertainty bands were widened by the increased noise standard deviation, shifting the mean trajectory. A larger recovery boost (Δ) provided a higher post-break level. However, it

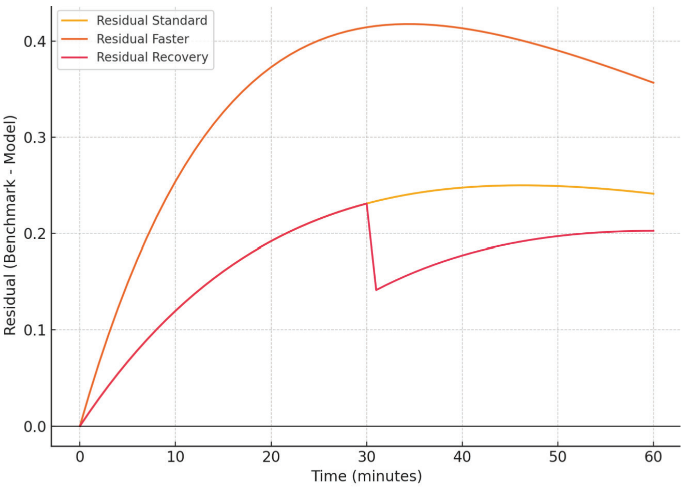

Figure 2. Residuals of Standard, Faster, and Recovery-Integrated Model Compared to Stroop Benchmark. Residuals-based summary of time-specific discrepancies between each model and the Stroop benchmark. Positive values showed over-prediction of accuracy in the model. Lines indicated mean residual, and shared regions were 95% pointwise bands from simulations. Increasingly positive residuals after around 20 minutes were shown in the faster model. In the recovery model, a localized deviation was shown around the break. However, it returned to the pre-break decay rate.

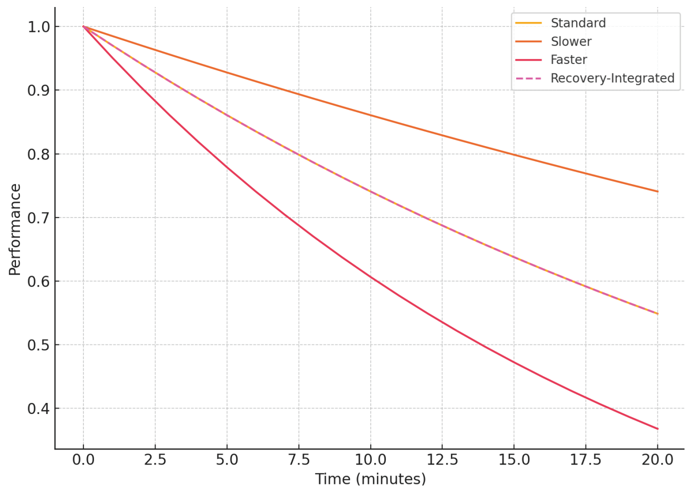

Figure 3. Performance Curves in the First 20 Minutes. Models were similar initially, then diverged. Trajectories for all four regimes and the benchmark were overlaid for the first 20 minutes. Early on, the differences in fatigue rate were minimal. For the faster model, it began to diverse within the first quarter hour. This indicated that meaningful differences in underlying fatigue dynamics may be masked by short tasks.

Table 2. Comparative Summary of Four Regimes Against Stroop Benchmark Over 60 Minutes

Model	Correlation (r)	Mean Squared Error	End Performance
Standard Fatigue	0.82	0.0062	0.15
Faster Fatigue	0.71	0.0117	<0.05
Slower Fatigue	0.89	0.0035	0.45
Recovery-Integrated	0.86	0.0043	~0.38

Pearson correlation (r), end-of-task performance (accuracy at minute 60), and mean square error were reported for four regimes: Standard, Faster, Slower, and Recovery-Integrated. A 1-minute rest was applied to the recovery regime with Δ = 0.08, while the same k was applied before and after the break. The better fit was indicated with higher r and lower mean square error.

left pre-break dynamics unchanged. Therefore, more performance was preserved by earlier breaks by 60 minutes than later ones (Figure 4).

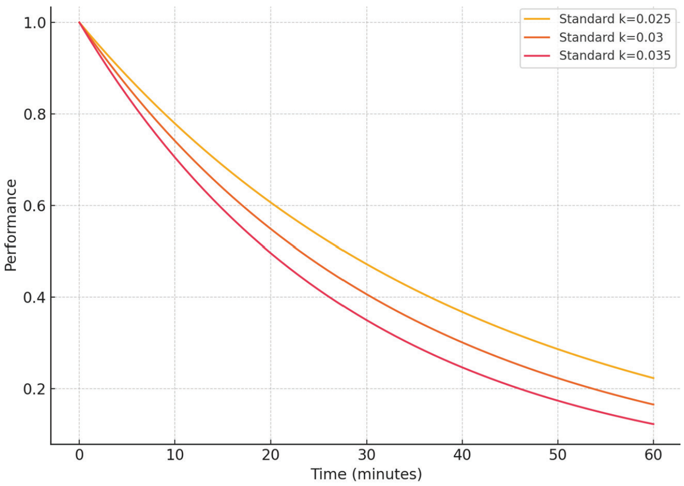

Figure 4. One-Factor Sensitivity Analysis Around the Fitted Standard Settings. Panels varied (i) fatigue rate k by {0.5, 0.75, 1.00, 1.25, 1.50} x standard, (ii) noise standard deviation by {0.015, 0.030, 0.045}, (iii) recovery boost Δ by {0.04, 0.08, 0.12}, and (iv) break timing by {20, 30, 40} minutes (recovery regime only). R = 10,000 simulations were run for each setting, calculating r, mean square error, and mean trajectories with 95% pointwise bands. Fit was degraded monotonically when moving k away from its fitted value. At the same time, larger noise widened bands without shifting the mean.

DISCUSSION

Meaningful insights are provided by the findings in this study about cognitive fatigue and the effectiveness of different quantitative models to capture performance decline over time. By comparing fatigue progress-based quantitative models and their results, it was shown that exponential decay models were differently suited to represent fatigue in the real world, and recovery mechanisms may improve ecological validity.

The results in the slower fatigue model as the best model with strong alignment with the benchmark Stroop task data through 60-minute task engagement supported the results in previous studies (13) indicating that cognitive performance gradually declines rather than sudden drop in sustained attention tasks. This result also implies that fatigue progression is a slow process in accumulation rather than an immediate drop.

The results in the standard fatigue model indicated an overestimate of fatigue effects later in the session. This implies that the model may not capture adaptive processes, such as strategic pacing or overall rest duration taken by participants during the task period. Even though the standard fatigue model fits well as a baseline, fatigue-related behaviors may not be fully captured by this model.

The results in the faster fatigue model indicated huge divergence from the benchmark Stroop task data, and hence did not represent cognitive fatigue in real-world settings. However, it supports the previous study (14) in a sense that it may be useful to simulate high-stress or sleep-deprived conditions with rapid drop in performance. With this model, the versatility of the modeling process is explained in different conditions.

The results in the recovery-integrated model showed an approximated benefit of a short break included in the task engagement. This temporary improvement from a short break in the simulated rest period supports the previous studies (12) about how a short break may restore attention and delay the fatigue onset. Inclusion of this model improves the robustness of the modeling framework mirroring the real-world work and learning environment, where productivity is usually maintained with the rest as a critical condition.

Finally, the robustness of models in this study was confirmed with the sensitivity analysis with small changes in parameters used in four different models. It was revealed that substantial differences may be introduced in predicted performance, especially for longer task durations, even with slight changes in the decay constant.

While this study provides important insight into cognitive fatigue and performance decline over time, there are limitations. First, neurobiological complexity related to fatigue, such as circadian rhythms and adaptive control strategies were not captured by the quantitative models in this study. Second, the framework in this study relied on a single benchmark – the Stroop task – sourced from a previous study. Even though Stroop was a well-known measure of sustained attention, it indicated only one form of cognitive engagement. This being said, other tasks may show different fatigue dynamics that were not specifically captured by the models in this study. Third, this study did not empirically validate prospective model predictions against independent human-subject data. Therefore, external validity of results still needs to be established. Fourth, a simplified single and mid-task break was assumed in the recovery-integrated regime to yield a one-time boost. However, in practice, recovery trajectories may likely be shaped by the number, length, and quality of breaks as well as inter-individual differences.

CONCLUSION

This study demonstrates a possibility of using quantitative models to explain cognitive fatigue in neural circuits in an extended period of task engagement. By applying standard, slower, faster, and recovery-integrated exponential decay models, this study shows that performance of tasks may be predicted by choosing different decay parameters. The slower fatigue model turned out to be the most accurate model, highlighting gradual accumulation of fatigue under typical task engagement conditions over time. The recovery-integrated model also indicated that even short breaks may significantly improve task performance and delay the fatigue onset through extended engagement.

The findings in this study confirmed the hypothesis that a slower exponential decay model and a recovery-integrated model may more accurately predict cognitive performance decline over time compared to standard or faster fatigue models. The results in this study support that the slower fatigue model represented the highest correlation and the lowest mean squared error against benchmark Stroop task data, and the recovery-integrated model showed improvement in task performance when short rest period was included. From the limitations of this study – with our models being only a foundational framework for studying cognitive fatigue and task performance decline over time – it is suggested for future studies to incorporate more datasets and more sophisticated dynamics with adaptive modeling to further improve predictive power and relevance.

CONFLICT OF INTEREST

The author declares no conflicts of interest related to this work.

REFERENCES

1. Billones R, Liwang JK, Butler K, Graves L, Saligan LN. Dissecting the fatigue experience: a scoping review of fatigue definitions, dimensions, and measures in non-oncologic medical conditions. *Brain Behav Immun Health*. 2021; 15: 100266. doi:10.1016/j.bbih.2021.100266.
2. Abd-Elfattah HM, Abdelazeim FH, Elshennawy S. Physical and cognitive consequences of fatigue: a review. *J Adv Res*. 2015; 6 (3): 351-358. doi:10.1016/j.jare.2015.01.011.
3. Matthews G, Desmond PA, Neubauer C, Hancock PA, editors. The Handbook of Operator Fatigue. *Aldershot (UK): Ashgate*; 2012.
4. MacLeod CM. Half a century of research on the Stroop effect: an integrative review. *Psychol Bull*. 1991; 109 (2): 163-203. doi:10.1037/0033-2909.109.2.163.
5. Stroop JR. Studies of interference in serial verbal reactions. *J Exp Psychol*. 1935; 18 (6): 643-662. doi:10.1037/h0054651.
6. Ackerman PL, editor. Cognitive Fatigue: Multidisciplinary Perspectives on Current Research and Future Applications. Washington (DC): American Psychological Association; 2011. https://doi.org/10.1037/12343-000
7. Wang Z, Chang Y, Schmeichel BJ, Garcia AA. The effects of mental fatigue on effort allocation: Modeling and estimation. *Psychol Rev*. 2022; 129 (6): 1457–1485. doi:10.1037/rev0000365.
8. Audiffren M, Giersch A, Burle B. Training willpower: reducing costs and valuing effort. *Front Neurosci*. 2022; 16: 834645. doi:10.3389/fnins.2022.834645.
9. Zeller R, Williamson A, Friswell R. The effect of sleep-need and time-on-task on driver fatigue. *Transp Res Part F Traffic Psychol Behav*. 2020; 74: 15–29. doi:10.1016/j.trf.2020.08.001.
10. Kovanović V, Gašević D, Dawson S, Joksimović S, Baker RS, Hatala M. Does time-on-task estimation matter? Implications on validity of learning analytics findings. *J Learn Anal*. 2015; 2 (3): 81-110. doi:10.18608/jla.2015.23.6.
11. Ackerman R, Thompson VA. Meta-Reasoning: Monitoring and Control of Thinking and Reasoning. *Trends Cogn Sci*. 2017; 21 (8): 607–17. doi:10.1016/j.tics.2017.05.004.
12. Van Dongen HPA, Dinges DF. Sleep, circadian rhythms, and psychomotor vigilance. *Clin Sports Med*. 2005; 24 (2): 237-249. doi:10.1016/j.csm.2004.12.007.
13. Lorist MM, Boksem MAS, Ridderinkhof KR. Impaired cognitive control and reduced cingulate activity during mental fatigue. *Brain Res Cogn Brain Res*. 2005; 24 (2): 199-205. doi:10.1016/j.cogbrainres.2005.01.018.
14. Caldwell JA, Caldwell JL, Schmidt RM. Fatigue countermeasures in aviation. *Aviat Space Environ Med*. 2009; 80 (1): 29-59. doi:10.3357/ASEM.2435.2009.