

## ORIGINAL ARTICLE

# Psychological well-being and multimodal predictors of operational safety in heavy-haul railway drivers

Jianhua Wang<sup>1,#</sup>, Yanming Ren<sup>2,#</sup>, Ting Meng<sup>1</sup>, Kaigong Zhao<sup>3</sup>, Shuyi Jia<sup>2</sup>, Xin Gao<sup>2</sup>, Yichen Huang<sup>2</sup>, Qi Gao<sup>2</sup>, Zhijie Liang<sup>2</sup>, Wei Li<sup>4</sup>, Pei Sun<sup>2,4,\*</sup>

<sup>1</sup>Jiliang Company, Guoneng Shuohuang Railway Development Co., Ltd., Cangzhou 061100, Hebei Province, China

<sup>2</sup>Faculty of Health and Wellness, City University of Macau, Macau 999078, China

<sup>3</sup>Safety and Environmental Protection Supervision Department, National Energy Investment Group Co., Ltd., Beijing 100011, China

<sup>4</sup>School of Social Sciences, Tsinghua University, Beijing 100084, China

## ABSTRACT

In high-risk, high-workload heavy-haul railway operations, locomotive crew safety performance depends not only on technical proficiency but is also strongly influenced by psychological and physiological states. Using 1117 real-world operational records from 203 heavy-haul railway drivers, this study examined the predictive roles of psychological health, cognitive task performance, and physiological indicators in distinguishing perfect from non-perfect operation events. The results indicated that better overall psychological health at the driver level was associated with a lower incidence of non-perfect operations. In contrast, performance on individual cognitive tasks did not show stable independent predictive effects. Physiological indicators were not significant overall but demonstrated a protective association under specific cognitive load conditions, particularly during digit memory tasks. Together, these findings suggest that operational safety is more closely related to drivers' general psychological health across tasks than to isolated cognitive ability measures. The results highlight the value of integrating psychological, cognitive, and physiological indicators when assessing operational risk among heavy-haul railway drivers.

**Key words:** heavy-haul railway, operational safety, psychological health, cognitive tasks, physiological, multimodal assessment

## INTRODUCTION

Heavy-haul railway transportation is a cornerstone of large-scale energy and bulk commodity distribution in China, operating within a safety-critical system characterized by high workload, strong system coupling, and extremely low tolerance for error (Perrow, 1984; Reason, 2000). In this context, locomotive drivers occupy a

central position in the human-machine-environment interaction chain and are required to sustain stable attention, judgment, and operational control over prolonged duty periods and under tightly constrained conditions (Fan et al., 2022; Wickens et al., 2004). Failures in this system carry substantial social and economic costs, extending beyond immediate operational disruption to broader supply-chain and public safety

<sup>#</sup>These authors contributed equally to this work.

### \*Corresponding Author:

Pei Sun, Faculty of Health and Wellness, City University of Macau, Avenida Padre Tomás Pereira, Taipa, Macau 999078, China. Email: peisun@cityu.edu.mo; <https://orcid.org/0000-0002-5938-8180>

Received: 19 December 2025; Revised: 24 December 2025; Accepted: 25 December 2025

<https://doi.org/10.54844/wsr.2025.1121>

 This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, which allows others to copy and redistribute the material in any medium or format non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

risks (Perrow, 1984).

Contemporary railway human factors research increasingly emphasizes that major safety incidents rarely result from isolated technical failures or simple "driver error". Instead, operational risk emerges from the interaction of individual states, organizational arrangements, and system design, ultimately manifesting at the level of human performance (Kusumastuti *et al.*, 2025; Reason, 1997). Accordingly, drivers' psychological functioning—shaped by fatigue, stress, workload, and emotional regulation—has become a central concern in safety management. Empirical studies consistently show that fatigue and psychological workload are closely linked to impaired performance and elevated risk among train drivers, while acute stressors may influence behavior even when drivers are not fully aware of their effects (Dorrian *et al.*, 2011).

Despite this growing recognition, important gaps remain in how driver wellbeing is conceptualized and measured in relation to operational safety. First, much of the existing literature relies on single-domain indicators, such as fatigue scales or isolated performance metrics, even though real-world safety is likely shaped by interacting psychological processes. From a well-being perspective, psychological health at work reflects a configuration of mental fatigue, perceived workload, self-efficacy, stress experience, and emotional state, rather than any single dimension in isolation (Demerouti *et al.*, 2001; World Health Organization, 2024). Second, the predictive value of laboratory-based cognitive tasks for real-world safety outcomes remains contested. Although such tasks provide standardized measures of cognitive ability, their ecological validity in complex, dynamic operational environments is limited, and their effects may be overshadowed by more stable, context-general psychological states (Chang *et al.*, 2012).

In response to these limitations, recent research has advocated multimodal approaches that integrate psychological, behavioral, and physiological indicators to improve risk detection and state assessment in safety-critical occupations (Li *et al.*, 2022). Evidence from driving and fatigue monitoring suggests that combining subjective reports with behavioral and physiological signals can enhance sensitivity to risk states. However, how these different modalities jointly contribute to safety outcomes—and whether their effects are hierarchical or context-dependent—remains insufficiently understood, particularly in highly professionalized groups such as heavy-haul railway drivers.

Against this background, the present study examines operational safety among heavy-haul railway drivers using a multimodal framework that integrates psychological health, cognitive task performance, and physiolo-

gical indicators. Operational performance is analyzed using an event-rate approach, treating non-perfect operations as risk events in line with safety management logic. The study addresses a focused theoretical question: In a population of well-trained drivers, are differences in safety outcomes better explained by task-specific cognitive abilities, or by more stable psychological health states together with context-dependent physiological conditions that generalize across operational contexts? By answering this question, the study aims to clarify the relative roles of psychological wellbeing, cognition, and physiology in safety prediction, and to inform the development of more effective, wellbeing-oriented risk monitoring systems in high-risk occupational settings.

## METHOD

### Participants

The study was conducted using data collected during routine operational monitoring, without any experimental manipulation or intervention. The research protocol was reviewed and approved by the institutional ethics committee (Approval No. THU-2022-23), and all participants were informed that psychological, physiological, and operational data collected during routine work activities could be used for research purposes; written informed consent was obtained prior to their inclusion in the study. Participants were heavy-haul locomotive drivers employed by a large railway transportation company in China. The final analytical sample consisted of 203 drivers, who together contributed 1117 operational records during the study period, with each record corresponding to a completed duty cycle with valid operational performance data and matched psychological or behavioral assessments.

Because drivers were assessed repeatedly during routine operations, the data had a multilevel structure, with operational records nested within drivers. Of the 1117 operational records, 91.1% represented full-score performance, whereas 99 records (8.9%) were classified as non-full-score events, indicating the occurrence of at least one operational deviation during the evaluated duty.

Psychological well-being measures were available for all operational records. Cognitive functioning was assessed using a set of computerized tasks administered on a rotating basis during routine operations; consequently, task-specific sample sizes varied. Across tasks, the number of drivers providing valid data ranged from 81 to 142, with corresponding numbers of operational records ranging from 150 to 241.

All participants were active-duty locomotive drivers at the time of data collection. Operational performance

data were obtained from the company's standardized assessment system. The study was based exclusively on data collected in routine work settings, without experimental manipulation.

## Measures

### *Psychological well-being questionnaire*

Psychological well-being was assessed using a self-developed mental health questionnaire tailored for heavy-haul locomotive drivers. The instrument was designed to balance brevity with multidimensional coverage and to capture psychological characteristics most relevant to locomotive operation. It comprises five dimensions: Mental fatigue, workload, self-efficacy, stress level, and emotional state.

Confirmatory factor analysis supported a five-factor structure with satisfactory model fit ( $\chi^2/df = 1.83$ , root mean squared error of approximation [RMSEA] = 0.056, Bentler's comparative fit index [CFI] = 0.94, Tucker-Lewis index [TLI] = 0.93, standardized root mean square residual [SRMR] = 0.048), which was considered acceptable given the brief and operationally embedded nature of the questionnaire. Standardized factor loadings ranged from 0.68 to 0.89. Composite reliability values ranged from 0.82 to 0.91, and average variance extracted values ranged from 0.56 to 0.73, indicating good convergent and discriminant validity. Cronbach's alpha coefficients for the total scale and subscales ranged from 0.75 to 0.88, demonstrating satisfactory internal consistency.

### *Cognitive tasks*

Cognitive functioning was assessed using a battery of computerized tasks administered on a rotating basis during routine operations. Each task targeted a distinct cognitive domain, and task-specific performance indicators were extracted for analysis. Because not all tasks were administered on every measurement occasion, sample sizes varied across tasks.

The Stroop task assessed executive control and inhibitory processing. Two indicators were extracted: Reaction time (in milliseconds [ms]), reflecting processing speed, and accuracy rate, reflecting response correctness.

The target tracking task assessed sustained attention and visuospatial monitoring ability. Performance was indexed by accuracy rate, reflecting the proportion of correctly tracked targets.

The Ligature test assessed visuomotor coordination and cognitive flexibility. Performance was indexed by total completion time (in seconds [s]), with shorter times indicating better task performance.

The digit memory task assessed short-term working memory capacity. Performance was indexed by the number of digits correctly recalled (count), with higher scores indicating better working memory.

The BallSport task assessed speed perception ability. Performance was indexed by reaction time (s), with shorter times indicating better perceptual processing speed.

The Balloon task assessed risk-taking behavior and response persistence. Performance was indexed by the number of unexploded balloons (count), with higher values indicating more conservative and controlled task performance.

The cognitive task performance was summarized using task-specific indicators reflecting accuracy, reaction time, or count-based outcomes, depending on the task. Descriptive statistics for each task are presented separately due to differences in measurement units and administration frequency.

### *Physiological indicator*

The physiological indicator used in this study refers to a composite physiological-behavioral index developed within the railway industry to assess drivers' physiological load and state stability during duty periods. Rather than representing a single biological mechanism, this indicator was designed to capture drivers' overall physiological readiness and regulatory capacity in operational contexts.

Physiological data were collected through an integrated multimodal assessment system deployed in routine operational settings. The indicator integrates multiple signal sources, including cardiorespiratory activity, fatigue-related physiological features, patterns of alertness fluctuation, and operational behavioral metrics. Specifically, the physiological indicator incorporates: (1) Cardiorespiratory features, such as heart rate and heart rate variability (HRV), which serve as indirect indicators of autonomic nervous system activity and physiological arousal; (2) fatigue-related physiological indicators, reflecting cumulative physiological strain and reduced recovery capacity; and (3) behavioral and psychophysiological features, extracted from operational behavior patterns and affective computing-based analyses of facial expressions, speech prosody, and eye-movement dynamics, which reflect attentional engagement, emotional arousal, and fatigue-related behavioral signatures.

All physiological and behavioral features were temporally aligned and standardized prior to analysis. Feature integration was implemented using a multimodal fusion framework, in which the relative contribution of

different signal sources was jointly modeled to generate an overall physiological state score. When multiple physiological records were available for the same driver within a given day, their mean value was used in subsequent analyses.

Higher values on this physiological indicator indicate better physiological regulation, lower fatigue levels, and greater readiness for safe operation, whereas lower values reflect increased physiological strain or potential fatigue-related risk. Similar composite physiological-behavioral indicators have been widely applied in safety-critical industries, such as aviation and railway operations, to support monitoring of alertness, fatigue, and operational risk (Hockey, 2013; Lim & Dinges, 2010; Borghini *et al.*, 2014). Moreover, prior research on multimodal integration of physiological and behavioral data has demonstrated the feasibility of using composite indices to estimate individuals' psychophysiological states based on multi-source signals (Hssayeni & Ghoraani, 2021). Accordingly, the physiological indicator adopted in the present study is consistent with established industry practices and current advances in multimodal physiological monitoring research.

#### *Operational performance*

Operational performance was measured using standardized operational scores recorded by the railway company. For inferential analyses, performance was operationalized as a binary indicator (full score *vs.* non-full score), reflecting whether any operational deviation occurred during the evaluated duty cycle.

In heavy-haul railway operations, performance evaluation follows a zero-tolerance safety logic: Any deviation from standard operating procedures—regardless of its magnitude—is treated as an operational deviation. Given the extreme train mass, long braking distances, and high systemic coupling, even minor deviations may substantially increase safety risks. Accordingly, a non-full score does not merely indicate lower performance, but represents the presence of a safety-relevant operational deviation. Distinguishing between full-score and non-full-score performance, therefore, constitutes an industry-consistent and safety-relevant categorization, rather than a coarse simplification of performance variability.

#### *Statistical analysis*

Data were collected between April and November as part of routine operational monitoring procedures. Statistical analyses were conducted in three steps. First, the main variables—psychological well-being, the physiological indicator, and the six cognitive task measures—were summarized using descriptive statistics to characterize their distributions. Continuous variables

were reported as means and standard deviations, with medians and interquartile ranges additionally provided when distributions showed noticeable skewness. Categorical variables were expressed as counts and percentages. Prior to regression analyses, all continuous predictors were standardized to z-scores to ensure comparability across measures with different units and ranges. Operational safety outcomes were expressed as event counts (non–full-score operations), with the total number of operations included as an offset to account for differences in work exposure across drivers.

Next, to examine how psychological well-being relates to operational safety at the driver level, we fitted Poisson event-rate models. For each driver, the number of non-full-score events was treated as the outcome, and the log of the total number of operations was included as an offset. This allowed us to estimate whether drivers with lower psychological well-being or poorer physiological status showed a higher rate of performance deviations after adjusting for differences in work exposure.

Finally, to test whether these associations held across different cognitive profiles, we conducted a set of joint Poisson models within each cognitive task subsample. Each model included the psychological well-being index, the corresponding cognitive task score, and the physiological indicator as predictors, again using the number of non-full-score events as the outcome and total operations as an offset. These models were used to evaluate the unique and combined contributions of psychological, cognitive, and physiological factors to operational performance.

All statistical analyses were performed using R (Version 4.4.3; R Foundation for Statistical Computing, Vienna, Austria) within the RStudio integrated development environment (Posit Software, PBC, Boston, MA, USA). The threshold for statistical significance was set at  $P < 0.05$ .

## RESULT

### ***Descriptive statistics and group comparisons***

A total of 1117 operational records from 203 drivers were included in the analyses, of which 91.1% achieved full-score performance. Descriptive statistics for psychological well-being dimensions, the overall psychological well-being composite, and the physiological indicator are presented in Table 1.

Among the psychological well-being dimensions, self-efficacy showed the highest mean value ( $7.18 \pm 1.89$ ), followed by mental fatigue ( $6.81 \pm 2.14$ ). In contrast, workload ( $3.09 \pm 1.93$ ) and stress level ( $3.92 \pm 2.41$ ) exhibited lower mean values. The score of overall

**Table 1: Descriptive statistics of psychological well-being dimensions and physiological indicator**

Variable	Mean	Standard deviation
Mental fatigue	6.81	2.14
Workload	3.09	1.93
Stress level	3.92	2.41
Self-efficacy	7.18	1.89
Emotional state	5.89	1.90
Overall psychological well-being	5.38	0.95
Physiological indicator	4.04	2.20

The overall psychological well-being score was calculated as the equally weighted mean of the five psychological dimensions (each weighted at 0.20). Higher scores indicate better psychological functioning after score transformation.

psychological well-being, calculated as an equally weighted composite of the five dimensions was  $5.38 \pm 0.95$ . The composite physiological indicator demonstrated moderate variability across records ( $4.04 \pm 2.20$ ).

Baseline performance across cognitive tasks showed a high degree of stability (Table 2). Group comparisons indicated that nearly all cognitive task measures did not differ significantly between the full-score and non-full-score groups ( $P > 0.05$ ), with mean values remaining highly similar across groups—for example, Stroop accuracy (0.800 *vs.* 0.807). Similarly, individual psychological dimensions showed no substantial group differences. However, the overall psychological well-being composite score was significantly lower in the non-full-score group ( $M = 5.16$ ) compared with the full-score group ( $M = 5.40$ ,  $P = 0.03$ ).

### **Poisson event-rate models**

At the driver level, 203 drivers contributed 1117 operational records, of which 99 were non-full-score events (event rate = 0.089 per operation). In a Poisson event-rate model with the logarithm of total operations as an offset, higher psychological well-being was significantly associated with a lower rate of non-full-score events (incidence rate ratio [IRR] = 0.80, 95% confidence interval [CI, 0.65, 0.99],  $P = 0.040$ ). In contrast, the physiological indicator was not significantly related to event rates at the driver level (IRR = 0.90, 95% CI [0.65, 1.22],  $P = 0.510$ ). No evidence of overdispersion was observed (dispersion ratio = 1.05,  $P = 0.295$ ), supporting the adequacy of the Poisson event-rate specification.

### **Joint effects of psychological well-being, cognitive tasks, and physiological indicators**

To examine whether psychological well-being remained associated with operational risk when cognitive performance and physiological status were considered jointly, a series of Poisson event-rate models was estimated. Each model included psychological well-being, one cognitive task, and a physiological indicator,

and was fitted to the subsample of drivers who completed the corresponding cognitive task.

Across all cognitive task models, psychological well-being tended to be associated with lower incidence rate ratios across subsamples (Table 3). Although these associations did not reach statistical significance in most task-specific models, the converging direction across tasks suggests a stable tendency rather than task-specific effects. Specifically, higher psychological well-being was associated with a lower rate of non-full-score operational events across diverse cognitive contexts.

In contrast, Individual cognitive task performance did not show a consistent independent association with event rates at the driver level. The incidence rate ratios for cognitive tasks varied in direction and were not statistically significant across models, suggesting that task-specific cognitive performance did not independently account for variations in operational risk when psychological well-being was included.

The physiological indicator likewise did not show consistent independent effects across tasks. However, a significant protective association emerged in the digit memory task subsample, where higher physiological scores were associated with a substantially lower event rate. This finding suggests that physiological state may play a buffering role under specific cognitive load conditions, rather than serving as a general risk predictor.

## **DISCUSSION**

The finding that the overall composite score reached statistical significance while individual dimensions did not supports the theoretical premise that operational safety is governed by a global regulatory state rather than isolated psychological stressors.

### **Overview of key findings**

The present study investigated whether operational safety among heavy-haul railway drivers is better

**Table 2: Descriptive statistics of cognitive task performance**

Task	N_trials	Mean	Standard deviation
Stroop-accuracy	241	0.801	0.141
Stroop-time (ms)	241	1182	391
Target tracking	196	0.740	0.216
Ligature test (s)	181	67.600	20.800
Digit memory	176	6.670	1.830
BallSport (s)	173	1.520	0.810
Ballon	150	10.200	3.490

Reaction time for the Stroop task is reported in milliseconds (ms), whereas reaction time for the BallSport task and completion time for the Ligature Test are reported in seconds (s). Accuracy-based indicators are expressed as proportions, and count-based indicators represent raw counts.

**Table 3: Joint effects of psychological well-being, individual cognitive tasks, and physiological indicators on the risk of non-full-score operational events (Poisson event-rate models)**

Cognitive task	Drivers (N)	Non-full events	Event rate	Psychological well-being IRR [95% CI]	P	Cognitive task IRR [95% CI]	P	Physiological IRR [95% CI]	P
Stroop-accuracy	142	23	0.095	0.88 [0.56, 1.38]	0.575	1.06 [0.67, 1.84]	0.807	1.19 [0.75, 1.77]	0.422
Stroop-time	142	23	0.095	0.89 [0.57, 1.40]	0.604	1.01 [0.63, 1.60]	0.963	1.20 [0.75, 1.80]	0.411
Target tracking	105	20	0.102	0.88 [0.54, 1.45]	0.602	1.05 [0.68, 1.68]	0.845	0.66 [0.38, 1.13]	0.139
Ligature-test	100	18	0.099	0.68 [0.40, 1.15]	0.160	0.90 [0.56, 1.55]	0.688	0.89 [0.47, 1.47]	0.679
Digit memory	99	14	0.080	0.90 [0.53, 1.48]	0.684	0.78 [0.44, 1.40]	0.409	0.26 [0.09, 0.68]	0.012
BallSports	94	15	0.087	0.73 [0.38, 1.34]	0.324	1.11 [0.67, 2.29]	0.731	1.07 [0.60, 1.82]	0.797
Balloon	81	9	0.060	0.54 [0.27, 1.09]	0.083	1.22 [0.59, 3.38]	0.646	1.20 [0.60, 2.26]	0.580

Each model includes psychological well-being, one cognitive task, and physiological indicator and is fitted on the subsample of drivers who completed the corresponding cognitive task. The outcome is the number of non-full-score events, with the logarithm of total operations used as an offset. IRR < 1 indicates a reduced rate of non-full-score events. All models were estimated using Poisson regression with an offset for the logarithm of total operations. For time-based cognitive measures (*e.g.*, Stroop time, BallSport, Ligature test), scores were reverse-coded so that higher values indicate better performance. IRR, incidence rate ratio; CI, confidence interval.

explained by task-specific cognitive abilities or by more general psychological and physiological states. Three core findings emerged.

First, overall psychological well-being showed a stable and protective association with operational safety, such that drivers with better psychological health exhibited lower rates of non-perfect operational events. Second, performance on individual cognitive tasks did not independently predict operational risk when psychological well-being was taken into account. Third, physiological indicators demonstrated context-dependent effects, becoming predictive only under specific cognitive load conditions, most notably in the digit memory task.

Together, these findings suggest that operational safety in this highly professionalized context is shaped less by isolated cognitive skills and more by cross-context psychological regulation, with physiological state exerting conditional influence under elevated cognitive demands.

### **Psychological well-being as a cross-context predictor of operational safety**

The most consistent result of the present study is the protective role of psychological well-being across operational contexts. Higher levels of overall psychological health were associated with lower rates of non-perfect operations, both at the driver level and across multiple task-specific subsamples.

Psychological well-being in this study represents a composite of mental fatigue, perceived workload, stress, emotional state, and self-efficacy (Bandura, 1997). Rather than capturing momentary performance capacity, this construct reflects drivers' sustained ability to regulate attention, emotion, and effort over prolonged duty periods (Ryff & Singer, 2008). In safety-critical environments such as heavy-haul railway operations, performance failures are rarely attributable to isolated cognitive lapses; instead, they emerge from cumulative strain, reduced self-regulatory capacity, and diminished resilience over time (Hockey, 2013; Matthews *et al.*, 2014).

This interpretation is consistent with human factors theories emphasizing system-level and cross-temporal influences on safety performance. From a Swiss cheese perspective, psychological well-being functions as a higher-order defensive layer that modulates how effectively lower-level cognitive and behavioral processes are deployed (Reason, 1997; Reason, 2000). Drivers with better psychological health are more likely to maintain stable vigilance, detect early signs of fatigue or overload, and adaptively regulate their behavior in response to operational demands (Hancock & Warm, 1989; Wickens *et al.*, 2004).

The findings also align with the broaden-and-build theory of positive psychological functioning, which posits that positive emotional and psychological states broaden attentional scope and build enduring regulatory resources (Fredrickson, 2001). In the present data, psychological well-being did not enhance performance on any single task per se; rather, it appeared to reduce the likelihood of safety-relevant deviations across diverse contexts, supporting its role as a general regulatory resource rather than a task-bound enhancer (Britt *et al.*, 2016).

### **Why do individual cognitive tasks not independently predict safety outcomes**

Contrary to expectations derived from traditional cognitive performance models, none of the six cognitive tasks independently predicted operational risk when psychological well-being was included in the models. This pattern does not imply that cognitive abilities are irrelevant for safe railway operation, but rather that their predictive value is constrained in this context.

First, laboratory-based cognitive tasks assess isolated cognitive components under controlled and time-limited conditions, whereas real-world railway operation requires sustained integration of attention, perception, decision-making, and emotional regulation under fatigue and stress (Matthews *et al.*, 2014; Wickens *et al.*, 2004). As such, the ecological validity of single-task cognitive measures for predicting long-term operational safety is inherently limited (Chang *et al.*, 2012).

Second, heavy-haul railway drivers constitute a highly selected and extensively trained professional group. Cognitive abilities in this population likely operate as qualification thresholds rather than differentiating factors, resulting in restricted variance. Well-established principles from personnel psychology indicate that under such conditions, even valid cognitive measures will show reduced predictive power for performance outcomes (Sackett *et al.*, 2008).

More fundamentally, cognitive capacity appears to be a necessary but not sufficient condition for operational

safety. The present results suggest that what matters most is not how well drivers perform on discrete cognitive tasks, but whether they can reliably mobilize their cognitive resources under conditions of fatigue, emotional fluctuation, and sustained workload (Hockey, 2013; Matthews *et al.*, 2014). Psychological well-being, rather than an isolated cognitive skill, governs this mobilization process.

### **Context-dependent role of physiological indicators**

Physiological indicators did not show a stable main effect across models, but a significant protective association emerged in the digit memory subsample. This pattern indicates that physiological state influences safety primarily under conditions of elevated cognitive demand.

The digit memory task places substantial demands on working memory and attentional control, functions that are particularly sensitive to fatigue and physiological strain (Baddeley, 2012; Lim & Dinges, 2010). Under such high-load conditions, variations in physiological regulation may directly translate into operational vulnerability. In contrast, under lower cognitive demand, drivers may compensate for suboptimal physiological states through effort or experience, masking physiological effects (Hockey, 2013).

This finding is consistent with resource vulnerability and compensatory control theories, which propose that physiological strain becomes behaviorally consequential when task demands approach or exceed available cognitive resources (Matthews *et al.*, 2014). It also aligns with prior fatigue and driving research showing that physiological indicators are more predictive of unsafe behavior during high-demand operational phases (Yu *et al.*, 2022; Zhang *et al.*, 2021).

Accordingly, physiological indicators should not be interpreted as global predictors of safety risk, but rather as conditional markers whose relevance depends on task demands and cognitive load.

### **Toward a hierarchical multimodal framework of operational safety**

Taken together, the present findings support a hierarchical multimodal framework for understanding operational safety in heavy-haul railway contexts. Psychological well-being emerged as the most stable, cross-context predictor, shaping how drivers sustain performance over time. Cognitive abilities, while essential for baseline competence, showed limited discriminative power within a highly trained population, consistent with range restriction effects (Sackett *et al.*, 2008). Physiological indicators contributed in a task-

dependent manner, becoming salient under conditions of heightened cognitive load.

This hierarchy suggests that effective safety monitoring should move beyond single-domain indicators toward integrated systems that prioritize psychological well-being as a foundational regulatory factor, while dynamically incorporating cognitive and physiological information in relation to operational context (Li et al., 2022; Reason, 1997). Such an approach may improve risk detection and support more targeted, wellbeing-oriented safety interventions in safety-critical industries.

## DECLARATION

### Acknowledgement

None.

### Author contributions

Jianhua Wang: Data curation, Project coordination; Yanming Ren: Conceptualization, Methodology, Writing—Original draft; Ting Meng: Data provision, Data validation; Kaigong Zhao: Data provision, Technical support; Shuyi Jia: Data collection, Data organization; Xin Gao: Data collection, Preliminary analysis; YiChen Huang: Data collection, Data cleaning; Qi Gao: Data collection, Data management; Zhijie Liang: Data collection, Administrative support; Wei Li: Supervision, Writing—Review and Editing; Pei Sun: Supervision, Conceptual guidance, Writing—Review and Editing.

### Source of funding

No funding.

### Ethical approval

This study was reviewed and approved by the institutional ethics committee (Approval No. THU-2022-23).

### Informed consent

This study involved human participants. Prior to participation, all individuals were fully informed about the study objectives and procedures, and written informed consent was obtained in accordance with ethical guidelines.

### Conflict of interest

Pei Sun is the Associate Editor-in-Chief of the journal. The article was subject to the journal's standard procedures, with peer review handled independently of the editor and the affiliated research groups.

### Use of large language models, AI and machine learning tools

The DeepSeek-V3.1-Terminus large language model was employed to assist with proof-reading of this

manuscript. All content remains the responsibility of the author.

### Data availability statement

Not applicable.

## REFERENCES

- Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1-29. <https://doi.org/10.1146/annurev-psych-120710-100422>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman and Company.
- Britt, T. W., Shen, W., Sinclair, R. R., Grossman, M. R., & Klieger, D. M. (2016). How much do we really know about employee resilience? *Industrial and Organizational Psychology*, *9*(2), 378-404. <https://doi.org/10.1017/iop.2015.107>
- Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, *44*, 58-75. <https://doi.org/10.1016/j.neubiorev.2012.10.003>
- Chang, Y. K., Labban, J. D., Gapin, J. I., & Etnier, J. L. (2012). The effects of acute exercise on cognitive performance: A meta-analysis. *Brain Research*, *1453*, 87-101. <https://doi.org/10.1016/j.brainres.2012.02.068>
- Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands-resources model of burnout. *Journal of Applied Psychology*, *86*(3), 499-512.
- Dorrian, J., Baulk, S. D., & Dawson, D. (2011). Work hours, workload, sleep and fatigue in Australian Rail Industry employees. *Applied Ergonomics*, *42*(2), 202-209. <https://doi.org/10.1016/j.apergo.2010.06.009>
- Fan, J., Smith, A. P., & Dinges, D. F. (2022). Fatigue, vigilance, and performance in train drivers: A systematic review. *Safety Science*, *149*, 105673. <https://doi.org/10.1016/j.ssci.2021.105673>
- Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist*, *56*(3), 218-226. <https://doi.org/10.1037/0003-066X.56.3.218>
- Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, *31*(5), 519-537. <https://doi.org/10.1177/001872088903100503>
- Hockey, G. (2013). *The psychology of fatigue: Work, effort and control*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139015394>
- Hssayeni, M. D., & Ghoraani, B. (2021). Multi-modal physiological data fusion for affect estimation using deep learning. *IEEE Access*, *9*, 21642-21652. <https://doi.org/10.1109/ACCESS.2021.3055933>
- Kusumastuti, D., Nicholson, A., & Fildes, B. (2025). Human reliability and organizational factors in railway safety management. *Safety Science*, *168*, 106308. <https://doi.org/10.1016/j.ssci.2024.106308>
- Li, W., Tan, R., Xing, Y., Li, G., Li, S., Zeng, G., Wang, P., Zhang, B., Su, X., Pi, D., Guo, G., & Cao, D. (2022). A multimodal psychological, physiological and behavioural dataset for human emotions in driving tasks. *Scientific Data*, *9*(1), 481. <https://doi.org/10.1038/s41597-022-01557-2>
- Lim, J., & Dinges, D. F. (2010). A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. *Psychological Bulletin*, *136*(3), 375-389. <https://doi.org/10.1037/a0018883>
- Matthews, G., Warm, J. S., Reinerman-Jones, L. E., & Langheim, L. K. (2014). Sustained attention and workload. *The Oxford handbook of cognitive engineering* (pp. 1-25). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199757183.013.017>
- Perrow, C. (1984). *Normal accidents: Living with high-risk technologies*. Princeton University Press.

- Reason, J. (1997). *Managing the risks of organizational accidents*. Ashgate.
- Reason, J. (2000). Human error: Models and management. *BMJ*, *320*(7237), 768-770. <https://doi.org/10.1136/bmj.320.7237.768>
- Ryff, C. D., & Singer, B. (2008). Know thyself and become what you are: A eudaimonic approach to psychological well-being. *Journal of Happiness Studies*, *9*(1), 13-39. <https://doi.org/10.1007/s10902-006-9019-0>
- Sackett, P. R., Lievens, F., Berry, C. M., & Landers, R. N. (2008). A cautionary note on the effects of range restriction on predictor validity. *Journal of Applied Psychology*, *93*(3), 538-544. <https://doi.org/10.1037/0021-9010.93.3.538>
- Wickens, C. D., Lee, J. D., Liu, Y., & Gordon-Becker, S. (2004). *An introduction to human factors engineering* (2nd ed.). Pearson.
- World Health Organization. (2024). *Guidelines on mental health at work*. World Health Organization.
- Yu, X., Wang, H., & Zhang, Y. (2022). Driver fatigue detection under high cognitive load using multimodal physiological signals. *IEEE Transactions on Intelligent Transportation Systems*, *23*(6), 5124-5135. <https://doi.org/10.1109/TITS.2021.3109876>
- Zhang, Y., Li, X., & Wang, J. (2021). Physiological indicators of driver fatigue under varying task demands. *Accident Analysis & Prevention*, *158*, 106203. <https://doi.org/10.1016/j.aap.2021.106203>