IETI Transactions on Ergonomics and Safety

http://ieti.net/TES/
2022, Volume 6, Issue 2, 15-30, DOI: 10.6722/TES.202212_6(2).0003.

TRAIN DRIVER UNDERLOAD AS A RESULT OF LOW TASK DEMAND AND MONOTONY

Baris Cogan[1, a], Birgit Milius[1,b]
[1]Technical University of Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany
[a]baris.cogan@tu-berlin.de, [b]birgit.milius@tu-berlin.de

Abstract Technological developments lead to substantial changes in train drivers' tasks. The effect of automation as well as changing task and environmental characteristics impose new challenges on train drivers. One of the expected effects is a rise in the underload of train drivers which in turn might affect their performance. In this study, scientific literature from various safety-critical domains is reviewed to identify the factors that have the potential to lead to cognitive underload. Experimental studies on train driving are extracted to conduct a simple meta-analysis on the effects of monotony and low task demand in train drivers. The findings point to impaired cognitive state in train drivers under monotonous and low task demand conditions. Prolonged performance under such conditions might impair vigilance task performance. Literature review reveals several good practices and potential techniques that can be used to avoid or delay the occurrence of underloading situations or to mitigate the arising effects.

Keywords: Mental workload; underload; automation; railway safety.

1. INTRODUCTION

In the future the main task of the train driver will be shifting from active controlling to passive monitoring [1]. The concept of grades of automation describes the main responsibilities of train drivers and automation systems [2]. For example, while in the GoA-1 train driver manually drives the train with the assistance of an automatic train protection, in the GoA-2 the speed control is automated, and the main tasks of the train driver is to monitor the tracks and the speed adherence. With the increase of automation levels, the main cognitive demands of train drivers become continuous routine information acquisition and processing. The effect of automation as well as changing task and environmental characteristics should be addressed for safe operation of rail systems. There is a large body of research on the effects of automation on human operators. It is expected that various aspects such as situation awareness, workload and complacency are negatively affected due to a changed cognitive state.

In the following, we focus on the relationship between workload and automation. There is no universally accepted definition of workload. However, the following definition contains widely accepted aspects of workload: *workload is not an inherent property, but rather it is determined by the interaction between the requirements of a task, the circumstances under which it is performed and the skills, behaviors and perceptions of the operator* [3].

Multiple Resources Theory states that people have different cognitive resources that they can use simultaneously to process task relevant information [4]. Mental workload reflects the level of attentional resources required to meet objective and subjective performance criteria, which is mediated by the characteristics of the task (e.g., demands), of the operator (e.g., skills), and the

http://ieti.net/TES/
2022, Volume 6, Issue 2, 15-30, DOI: 10.6722/TES.202212_6(2).0003.

environmental context [5]. The Malleable Attentional Resources Theory (MART) suggest that in a situation with low task demand, the mental capacity of the operator decreases to meet the demand level of the task [6]. Therefore, when a sudden increase in the task demand occurs (i.e., response to an unexpected event), the operator fails to cope with the issue.

A model of cognitive task load includes three main load factors: the percentage of time occupied, the level of information processing (i.e., the percentage knowledge-based actions), and the number of task-set switches [7].

There are also efforts to distinguish between task load and workload. A study in air traffic control domain states the distinction between -task load- the objective demand of a task and -workload- the subjective demand experienced by the operator performing the task [8].

On the other hand, workload alone is not a sufficient indicator of performance. For example, a train simulator study was unable to detect any differences in objective performance measures (i.e., speed control and response to critical events) despite the two different workload and fatigue levels reported by the participants [9]. For this reason, mediating effects and performance consequences of underloading situations need to be investigated. For example, prolonged and high-workload task conditions might generate active fatigue and continuously low workload situations might lead to passive fatigue [10]. A car driving study found that monotony could lead to fatigue symptoms which may cause vigilance performance decrements [11].

2. FACTORS ASSOCIATED WITH UNDERLOAD

To find the most relevant literature on workload and factors associated with it, several key search terms were searched in academic databases such as Web of Science, ScienceDirect, PubMed, Sage Journals and Google Scholar. Some of the search terms and their related terms include Workload-Underload, Time-on-task; Vigilance-Attention; Fatigue-Passive fatigue; Automation-Performance decrement. Domain-breakdown of the publications include rail, road, and air transport as well as other safety-critical domains and domain-independent literature such as theoretical studies on cognitive psychology. The differences between the domains need to be considered for a more detailed analysis as they can have an influence on workload.

The factors that are associated with workload, particularly with underload, are identified for the rail domain (see Table 1). The factors are presented under three categories based on how they influence the performance. These are the factors that are related to the task characteristics, the external factors that are related to the environment the tasks are being performed in, and the personal factors that influence the performance individually. These factors are extracted from experimental studies and qualitative studies of observations, surveys, or interviews. Additionally, factors that are adapted from car driving studies are indicated with an asterisk symbol (*).

http://ieti.net/TES/
2022, Volume 6, Issue 2, 15-30, DOI: 10.6722/TES.202212_6(2).0003.

Table 1. Factors that are associated with underload in railway domain.

Task-related Factors	Situational (external) factors	Individual factors
Monotony (task) [12]	In-cab environment*	Personality traits [13]
Objective task demand [14]	Driver-machine-interface design [15]	Route/rollingstock knowledge [15]
Time-on-task [12]	Weather/visibility [16]	Motivation* [17]
	Time of the day [18]	Experience [19]
	Monotony (environment) [20]	

3. UNDERLOAD AS A RESULT OF LOW TASK DEMAND AND MONOTONY

The present paper mainly focuses on objective task demand and monotony as defining aspects of workload. Objective demand of tasks is the part of the demand imposed on the operator by the task itself. Interface demand and procedure demand are some of the factors that contribute to the task load [8]. According to this view, the intensity of the task (e.g., the number of tasks) also influences the work demand operators must handle.

There is a large amount of overlap in the literature of monotony, fatigue, and vigilance, with all of them associated with the states of reduced arousal and alertness. Monotony stems from external factors such as repetitive environment or occurs due to the characteristics of a task such as the level of simplicity [12]. For example, characteristics of vigilance tasks are usually in the form of prolonged and continuous tasks or tasks with critical signals occurring infrequently and irregularly. Vigilance studies across different domains show that negative effects of time-on-task manifest itself after a 20-30 minutes of task performance [21]. Time-on-task effects may occur relatively sooner under particular conditions such as monotonous environment [12]. Although simple vigilance tasks are typically monotonous, monotonous tasks cannot always be classified as vigilance tasks [22]. However, some researchers distinguish the subjective experience of monotony without specifying the nature of the task.

The theoretical derived factors of monotony as stated above can also be found in studies with train drivers. There are several survey-based studies conducted on train drivers to obtain information on their ability to identify underload and its inducing factors, and countermeasures they use to overcome the effects of underload. A workshop with train drivers in the UK revealed that half of the drivers think they can realize when they are drifting into a state of underload [23]. A survey with 143 Australian train drivers showed that more than a third of drivers experience boredom or monotony on at least half the shifts they work [12]. Over half the drivers reported that this was more likely to occur when they drive the same route a few times in a row. Early morning shifts and sleep deprivation were also found to be factors influencing the experience of boredom and monotony.

In order to analyze the effect of the task load and the environment on monotony, the factors considered are limited to those given in Table 2. There are two reasons for this limitation. Firstly, in line with the literature review, the factors that could potentially lead to underload are considered under two categories, namely environmental and task related factors. Secondly, combining multiple factors with different characteristics might compromise the reliability of the meta-analysis. Although there are other factors, such as early morning shifts, that could lead to similar performance decrements, there

http://ieti.net/TES/
2022, Volume 6, Issue 2, 15-30, DOI: 10.6722/TES.202212_6(2).0003.

is a limited number of experimental studies that investigate the effect of such factors. Therefore, the term monotony in this paper will be used to refer to the situations where one or a combination of factors below are present.

Table 2. Factors that are considered in the meta-analysis.

Environmental factors	Task related factors
Unchanging environment	Lower task demand (i.e., reduced number of tasks due to automation)
Environment that changes in repetitive and predictive way	Change in task characteristics (i.e., passive monitoring)

4. META-ANALYSIS METHOD

This paper provides a literature review and a simple meta-analysis on the effects of underload and monotony on train driver cognitive state and performance.

The PICO model was selected to determine the search criteria [24]. PICO stands for Problem/Population, Intervention, Comparison and Outcome. It was operationalized as follows:

- Problem/Population: Rail Operator OR Train driver
- Intervention: Automation OR Task Demand OR Assistance OR Monotony
- Comparison: Driving OR Manual OR Workload
- Outcome: Workload OR Situation Awareness OR Performance OR Vigilance

Total records found were 251 in Web of Science, 177 in PubMed and 25 in additional sources such as citation searching. After the filtered search, the duplicate records were eliminated. Remaining records were screened through titles and keywords and, secondly, through abstracts. In the next step, study methodology and conclusion chapters of the remaining records were screened. The records were filtered based on following inclusion criteria to conduct a meta-analysis on experimental train driving studies;

- Railway domain (train driver)
- Experimental study (simulator or naturalistic)
- Level of automation or task demand is manipulated, or monotony is induced
- Measurements of driving performance and/or operator`s cognitive state are collected

13 studies were included in the first meta-analysis step. Four of these studies were later excluded either due to incomplete data or unsuitable variables. The final list of studies with the reference numbers can be found in the Table 4 (Annex-1). Literature was reviewed to find out if a meta-analysis in this area had already been conducted. A study conducted a meta-analysis of human- automation interaction studies focusing on workload and vigilance task performance [22]. The analysis considered the studies with a manual only group and the analysis were made by comparing the automation against the manual only group. Unlike the present study, the analysis was not limited to

http://ieti.net/TES/
2022, Volume 6, Issue 2, 15-30, DOI: 10.6722/TES.202212_6(2).0003.

any domain. Employing an analysis only on train driver studies decreases the differences between task characteristics amongst studies.

A data table with relevant information from each study was created. The shortened version of this table includes study information and characteristics such as publication year, study design, number of subjects, dependent and independent variables (See Table 4 in the Annex 1). Additionally, a simple task analysis was employed to group the independent variables (task demand, level of automation or monotony) within each study. Due to the limited information provided by each paper, only a conceptual task analysis could be conducted. Alternatively, a task factor rating method to estimate objective task demand and mental workload levels as suggested in another paper was considered [22]. However, this method is not suitable here, because it is not possible to collect the information on the number of executed sub-tasks during manual and automation levels in each study.

Underloading conditions in this analysis differ in two ways; task and environment related factors (Table 2). Environmental factors include monotonous environment (i.e., unchanging landscape), and task related factors could be related to objective task demand (i.e., automation aids or additional secondary tasks). For this reason, the relation between automation levels and the objective task demand should be addressed. It was proposed that automation can be applied to four broad classes of functions: information acquisition, information analysis, action selection and action implementation [25]. A meta-analysis study applied this model of automation stages to define relative automation levels of two systems [26]. According to this approach, higher automation can be represented by the total number of stages with higher levels, or higher levels in later stages. A similar approach was used in this study to group the task demand levels of studies that manipulated either the automation levels or task demand levels. The tasks of the human operator and the automated system were identified using the information given in each study. Besides the number of tasks that human operator executes, the stages of information processing for each level were used to ordinally rank these levels relative to each other. The system with the higher number of manual tasks imposes greater task demand on the operator, and the system that has the highest automation level imposes lowest demand on the operator. Thus, the lowest task demand creates a task-related underload condition. Analysis levels were coded as M1, M2, M3 and M4 in the order of increased underload (or monotony). The level M4 exists in three studies and represents the Autopilot level (monitoring-only) with vigilance tasks. This methodology was applied to all studies in the same manner to enable comparison between groups in different studies (i.e., experiment groups with similar task numbers and characteristics). However, as a limitation it should be noted that this ordinal presentation of task underload does not constitute the absolute levels of task demand. This is due to the differences in task characteristics and examined system of each study.

Performance measures and measurement methods of each study are identified. Performance measures with similar properties are grouped into four categories as dependent variables. Secondary task performance refers to tasks with vigilance characteristics such as response time to critical events. Parameters related to operator`s cognitive and perceptual state are grouped together. Finally, subjectively measured workload results are combined in one category. Only one study included physiological workload measurements, which is explained further later in the results section.

http://ieti.net/TES/
2022, Volume 6, Issue 2, 15-30, DOI: 10.6722/TES.202212_6(2).0003.

The following list contains the three categories of dependent variables;

- Secondary task performance: Reaction time to failures or critical events, reaction distance (i.e., distance to the point of the safety-critical event at the time of the action), accuracy of responses.
- Operator state: Situation awareness, fatigue, sleepiness, boredom, arousal.
- Mental workload: subjectively reported mental workload.

The approach used in this paper for data aggregation differs from classic meta-analysis methods with an effect size analysis (e.g., Hedge`s g or Cohen`s d). This is due to the lack of information in most of the studies in terms of effect size and other statistical data. The data aggregation method of three earlier studies are adapted to our analysis [22, 26, 27]. However, the way each variable is weighted differs from these studies.

A meta-analysis on the effect of automation on mental workload employed a scoring method to assign the mental workload outcomes (such as Manual>Automated) [22]. However, the study used only two comparison groups, and a rank order was set according to the mental workload outcome (i.e., higher workload in automated condition receiving the highest score). Additionally, if multiple measures of mental workload were used, the outcome was assigned to the majority finding or only one type of measure according to a priority order set by the authors. In our analysis, each comparison condition received a performance score, and multiple measures of a dependent variable are integrated as one overall outcome. This approach was used in a meta-analysis of the effects of automation on human performance earlier by researchers [26]. The performance ranking was made based on the lowest performing group, which always gets the score 0. Other groups then are scored based on their performance relative to each other (0 to 2). When the groups showed no performance differences, the same ranks were assigned to these groups.

If a dependent variable group was measured by more than one variable, the rankings of individual variables are integrated as one overall ranking. For example, if the subjective mental workload was measured by two different methods (e.g., NASA-Taskload Index and Subjective Work Underload Checklist), these measurements were combined as one overall mental workload result. On the other hand, when the mental workload is measured by both subjective and objective measurement tools (e.g., NASA-TLX and heart rate), these measurements are combined as one overall mental workload ranking. This condition happened only in one study. If these two measures conflict with each other, the overall rank is considered as zero effect, meaning no change is observed between these groups. For example, this situation occurred in study number 10 [28]. For the two driving conditions, measurements of objective and subjective situation awareness were conflicting. As both measurements failed statistical significance, the overall result was no-change in situation awareness. Table 3 below shows two examples of how the conditions were ranked based on the performance scores.

http://ieti.net/TES/
2022, Volume 6, Issue 2, 15-30, DOI: 10.6722/TES.202212_6(2).0003.

Table 3. An example of how the performance rankings of experiment groups were determined.

Study Nr	Experiment groups	Performance variables	Meta group	Measurements	Results	St. sign. (- or +)	Overall ranking
10	M2, M3	Situation awareness	Opertor state	Objective and subjective	Objective: M3>M2 Subjective: M2>M3	Objective: (-) Subjective: (-)	M2=M3
8	M1, M2	Mental workload	Workload	Subjective	M1>M2	(+)	M1>M2

5. RESULTS

Performance results of different experimental groups are presented as performance ranking for each category (see Table 3 for an example). Using these rankings, the performance change within meta groups is visualized on a graph (Figure 1 to 3, left).

Additionally, the slopes of these individual lines are combined in order to obtain the performance changes with the change in underload levels (Figure 1-3, right). Three statistical significance levels were defined; namely, statistically significant, partially significant, not significant. The results with multiple measurements that had at least one significant measurement were coded as partially significant. Based on the significance levels, the slope of each condition change (e.g., M1 to M2) is weighted and combined in a single slope. The slope line always starts with the value of zero and indicates the change in performance score with the change in monotony or task demand.

Figure 1 below shows the workload scores of five studies for each individual study. All studies except the study number 5.2 measured the subjective workload. The line for the study number 5.2 shows the combination of subjectively and physiologically measured workload. According to the results of this study, self-reported mental workload levels were not different amongst all groups.

There is an overall trend of decreasing workload scores as seen in the graphs. In the study with the only exception (7.2), the group that uses a driver assistance system reported higher workload levels. The DAS used in the study provides speed advice or timetable information, which are related to information acquisition and analysis stages. Participants reported that additional information sources made the ride more mentally demanding. The graph on the right shows the combined slopes of workload scores from individual studies. The overall trend of decreased workload resulting from the increase in monotony can be seen in the graphs.

Figure 2 represents the change in operator perceptual or cognitive state for various levels of monotony. Half of all studies reported statistically significance decrease in performance score under increasing underload. All three studies that recorded no difference failed to reach significance. Therefore, the overall trend indicates that monotonous conditions could lead to impairment in operators` cognitive or perceptual state.

IETI Transactions on Ergonomics and Safety

http://ieti.net/TES/
2022, Volume 6, Issue 2, 15-30, DOI: 10.6722/TES.202212_6(2).0003.

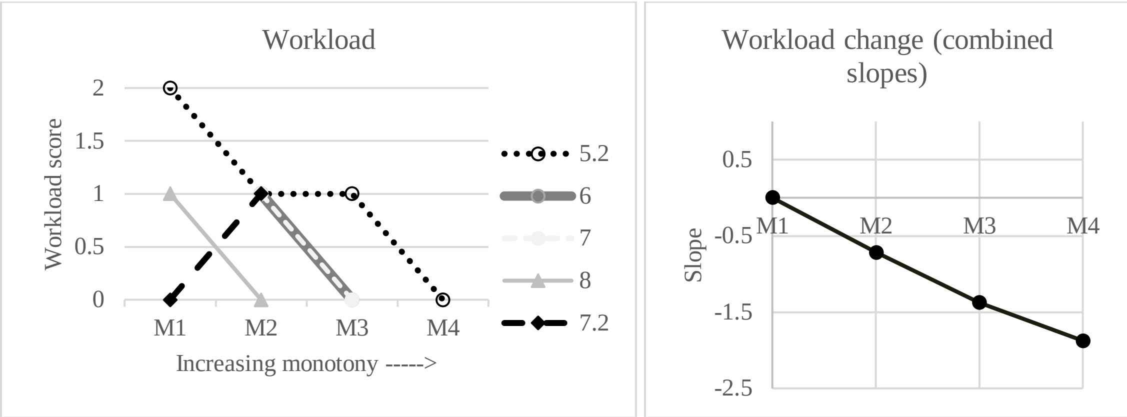


Figure 1. (left) Change in workload scores in individual studies, (right) Change in workload through weighted average of slopes of each study.

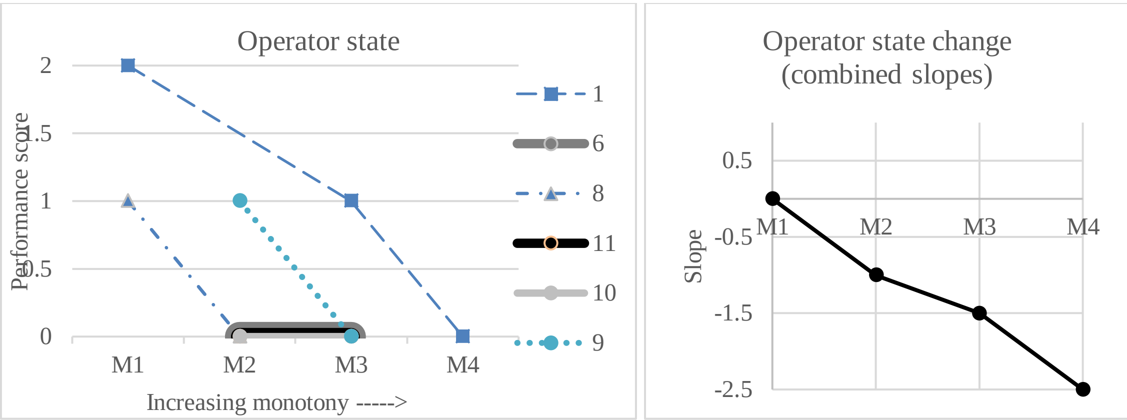
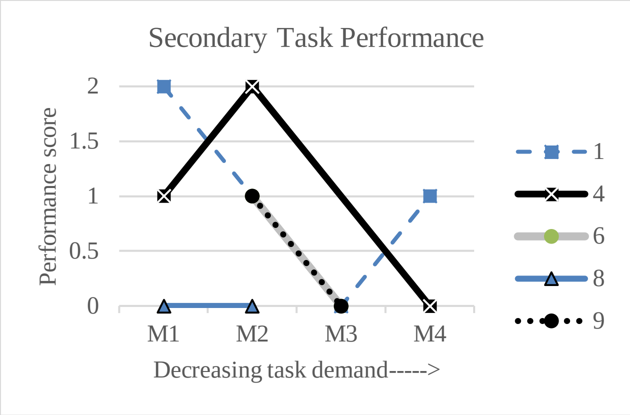


Figure 2. (left) Change in operator cognitive or perceptual state in individual studies, (right) Change in operator cognitive or perceptual state through weighted average of slopes of each study.

Secondary task performance involves vigilance tasks such as response time or distance to critical events as well as the accuracy of these responses (figure 3). In study number 5.2, the secondary task performance is excluded from the analysis [22]. According to the results of study 5.2, participants reported high subjective workload with no difference between the experiment groups (not significant). However, when combined with the significant results of SWUC ratings and heart rate records, the results suggest that the participants in the group M1 experienced overload. Therefore, the change in the performance could be the direct result of the shift from overload to moderate levels of workload. Authors of the study 5.2 argue that the participants experience underload in autopilot with a classical sensory vigilance task but report high subjective workload in NASA-TLX due to the state related effort (i.e., trying to stay alert) during extra-low levels of task demand. The graph in Figure 3 with the combined slopes of secondary task performance shows the performance decrement is seen after moderate levels of cognitive load and remains at similar levels with the further drop in cognitive load.

http://ieti.net/TES/
2022, Volume 6, Issue 2, 15-30, DOI: 10.6722/TES.202212_6(2).0003.



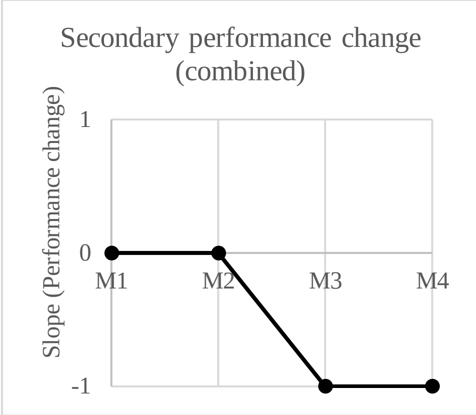


Figure 3. (left) Change in secondary (vigilance) performance in individual studies, (right) Change in secondary (vigilance) performance through weighted average of slopes of each study.

6. COUNTERMEASURES

Literature review and the meta-analysis suggest that low task demand and monotony might lead to underload and other cognitive decrements in human operators. Performance consequences of these conditions might primarily reveal themselves in vigilance tasks that require sustained attention and high situation awareness. Considering the task characteristics of train drivers, the main challenge is the detection of and decision making for the non-routine time-critical events. Scientific literature from different domains is reviewed to identify the most common countermeasures for preventing underload or mitigating its negative effects. Good practices or potential countermeasures that can be transferred to rail transport are identified. Key responsibilities of train drivers associated human factors issues and potential mitigation techniques are presented in Table 5 of Annex-1.

7. CONCLUSION

There is a large body of research on the general effects of increasing automation on operator and system performance. Routine system performance generally increases with the implementation of automation due to the reliable and efficient execution of various tasks. However, the decrement in failure performance emerges as the main issue in systems such as in intermediate automation where the human operator performance becomes crucial. In addition to the changes in the number or characteristics of the tasks due to automation, monotonous environment and prolonged execution of these tasks might cause vigilance performance decrements. This is particularly important for train drivers since the nature of their task and the environment is often monotonous.

The effects of lowered physiological activity under prolonged low task demand conditions were presented in earlier chapters. Decrements in cognitive abilities may lead to safety risks when these abilities are needed the most, e.g., to respond quickly to an unforeseen event. This overall effect was seen in the analysis conducted above. With the train driver`s role becoming more passive observer and supervisory, attentional problems may be a bigger issue for the drivers in unexpected or infrequent events than it was before.

The literature review reveals several good practices and other potential techniques that can be used to avoid or delay the occurrence of underloading situations or to mitigate the arising effects. Despite

http://ieti.net/TES/
2022, Volume 6, Issue 2, 15-30, DOI: 10.6722/TES.202212_6(2).0003.

most of the countermeasures being tested in specific situations, it could still provide an insight into potential techniques to overcome underload for train drivers. More empirical research in the context of train driving representing real situations needs to be conducted in order to come up with more concrete solutions.

8. LIMITATIONS

A limitation of this research is the relatively small number of studies included in this analysis. Including more experiments from other safety-critical domains would increase the size of the dataset, but at the same time, increase the differences between studies in terms of task characteristics. Some of the differences that already exist are task duration or number of task switches.

Performance scores are assigned to each experiment condition to visualize the performance differences between different task demand levels within a study. Using an ordinal ranking system to quantify the task demand and performance data makes it difficult to derive conclusive statistical results. However, this approach allows for scaling task demands based on limited information. On the other hand, this method of scaling might lead to over- or underestimation of trends due to weighted averages based on statistical significance.

Nevertheless, the structured review of literature combined with the overall patterns obtained from the analysis provide an extensive summary of knowledge on cognitive underload and its mediating effects on train driver performance.

Acknowledgements

The authors declare that there is no conflict of interest.

References

[1] Brandenburger, N., 2016, Is the driver´s traditional outside view still necessary to ensure situation awareness in high speed/ automated train operation?, In *Abstracts of the 58th Conference of Experimental Psychologists*, (p. 35), Heidelberg, Germany.

[2] UITP, 2018, World report on metro automation, Observatory of Automated Metros, Accessed 1 June 2021, https://www.uitp.org/publications/world-report-on-metro-automation/.

[3] Hancock, P. A., & Meshkati, N. (Eds.), 1988, *Human mental workload*, (Vol. 52), Elsevier.

[4] Wickens, C. D., 2008, Multiple resources and mental workload, *Human factors,* 50(3), pp. 449–455.

[5] Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A., 2015, State of science: mental workload in ergonomics, *Ergonomics,* 58(1), pp. 1–17.

[6] Young, M. S., & Stanton, N. A., 2002, Malleable attentional resources theory: a new explanation for the effects of mental underload on performance, *Human factors,* 44(3), pp. 365–375.

[7] Neerincx, M., 2003, Cognitive Task Load Analysis, In E. Hollnagel (Ed.), *Handbook of Cognitive Task Design*, (Vol. 20031153, pp. 283–305, Human Factors and Ergonomics), CRC Press,

[8] Hilburn, B.,Jorna, P. G. A. M., 2001, Workload and air traffic control., In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload and fatigue: theory, research and practice*, (p. 384).

[9] Robinson, D., Waters, S., Basacik, D., Whitmore, A., & and Reed, N., 2015, *A pilot study of low workload in train drivers*, TRL-RSSB Joint Research Board, TRL Limited.

[10] Desmond, P. A., Hancock, P. A., & Monette, J. L., 1998, Fatigue and Automation-Induced Impairments in Simulated Driving Performance, *Transportation Research Record: Journal of the Transportation Research Board,* 1628(1), pp. 8–14.

[11] Farahmand, B., & Boroujerdian, A. M., 2018, Effect of road geometry on driver fatigue in monotonous environments: A simulator study, *Transportation Research Part F: Traffic Psychology and Behaviour,* 58, pp. 640–651.

[12] Dunn, N. J., 2011, *Monotony: The Effect of Task Demand on Subjective Experience and Performance*, University of New South Wales, Sydney.

[13] Giesemann, S., & Naumann, A. B., Zugsicherungssysteme - Assistenz für Triebfahrzeugführer?, *Kognitive Systeme,* 1.

[14] Dunn, N. J., & Williamson, A., 2011, Monotony in the rail industry: The role of task demand in mitigating monotony-related effects on performance, In R. Mitchell (Ed.) *HFESA 47th Annual Conference*, Human Factors and Ergonomics Society of Australia,

[15] Verstappen, V. J., Marit S. Wilms, & van der Weide, R., 2017, The impact of innovative devices in the train cab on train driver workload and distraction, In *Sixth International Human Factors Conference*, (pp. 10–19), London, UK.

[16] van der Weide, R., Bruijn, D. de, & Zeilstra, M., 2017, ERTMS pilot in the Netherlands – impact on the train driver, In *Sixth International Human Factors Conference*, London, UK.

[17] Saxby, D. J., Matthews, G., Hitchcock, E. M., Warm, J. S., Funke, G. J., & Gantzer, T., 2008, Effect of Active and Passive Fatigue on Performance Using a Driving Simulator, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 52(21), pp. 1751–1755.

[18] Nurul Izzah Abdrahman, & Siti Zawiah Md Dawal, 2016, The mental workload and alertness levels of train drivers under simulated conditions based on electroencephalogram signals, *Malaysian Journal of Public Health Medicine 2016,* 16, pp. 115–123.

[19] Lo, J. C., Sehic, E., & Meijer, S. A., 2014, Explicit or Implicit Situation Awareness? Situation Awareness Measurements of Train Traffic Controllers in a Monitoring Mode, In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, A. Kobsa, F. Mattern, et al. (Eds.), *Engineering Psychology and Cognitive Ergonomics*, (Vol. 8532, pp. 511–521, Lecture Notes in Computer Science), Cham Springer International Publishing,

[20] Stein, J., & Naumann, A., 2016, Monotony, Fatigue and Microsleeps - Train driver' daily routine: a Simulator study, *Rail Human Factors. Proceedings of the 2nd German Workshop on Rail Human Factors, Milius, B; Naumann, A (Eds.)*, pp. 96–102.

[21] Edkins, G. D., & Pollock, C. M., 1997, The influence of sustained attention on Railway accidents, *Accident Analysis & Prevention,* 29(4), pp. 533–539.

[22] Spring, P. A., 2011, *Levels of Train Automation: Classification and Determining their Impact on Driver Mental Workload and Vigilance Task Performance*, PhD, University of New South Wales.

[23] Waters, S., 2019, *Identifying and evaluating techniques to mitigate cognitive underload for train drivers*, Rail Safety and Standards Board Limited, London, UK.

[24] da Costa Santos, C. M., Mattos Pimenta, C. A. de, & Nobre, M. R. C., 2007, The PICO strategy for the research question construction and evidence search, *Revista latino-americana de enfermagem,* 15(3), pp. 508–511.

[25] Parasuraman, R., Sheridan, T. B., & Wickens, C. D., 2000, A model for types and levels of human interaction with automation, *IEEE transactions on systems, man, and cybernetics. Part A, Systems and humans: a publication of the IEEE Systems, Man, and Cybernetics Society,* 30(3), pp. 286–297.

[26] Onnasch, L., Wickens, C. D., Li, H., & Manzey, D., 2014, Human performance consequences of stages and levels of automation: an integrated meta-analysis, *Human factors,* 56(3), pp. 476–488.

[27] Wickens, C. D., Li, H., Santamaria, A., Sebok, A., & Sarter, N. B., 2010, Stages and Levels of Automation: An Integrated Meta-analysis, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 54(4), pp. 389–393.

http://ieti.net/TES/
2022, Volume 6, Issue 2, 15-30, DOI: 10.6722/TES.202212_6(2).0003.

[28] Brandenburger, N., Stamer, M., & Naumann, A., 2016, Comparing different types of the track side view in high speed train driving, In *Human Factors and Ergonomics Society Annual Meeting Proceedings*. .

[29] Lanzilotta, E. J., & Sheridan, T. B., 2005, *Human Factors Phase III:* Effects of Train Control Technology on Operator Performance, U.S. Department of Transportation.

[30] Spring, P., Mcintosh, A., Caponecchia, C., & Baysari, M., 2012, Level of Automation: Effects on Train Driver Vigilance, *Rail Human Factors Around the World: Impacts on and of People for Successful Rail Operations*.

[31] Large, D., Golightly, D., & Taylor, E., 2014, The effect of driver advisory systems on train driver workload and performance, *Conference: Contemporary Ergonomics and Human Factors*, pp. 335–342.

[32] Brandenburger, N., Wittkowski, M., & Naumann, A., 2017, Countering Train Driver Fatigue in Automatic Train Operation, In *Sixth International Human Factors Conference*, London, UK.

[33] Beller, J., Heesen, M., & Vollrath, M., 2013, Improving the driver-automation interaction: an approach using automation uncertainty, *Human factors,* 55(6), pp. 1130–1141.

[34] Robertson, I. H., Tegnér, R., Tham, K., Lo, A., & Nimmo-Smith, I., 1995, Sustained attention training for unilateral neglect: theoretical and rehabilitation implications, *Journal of clinical and experimental neuropsychology,* 17(3), pp. 416–430.

[35] Wickens, C., 2018, Automation Stages & Levels, 20 Years After, *Journal of Cognitive Engineering and Decision Making,* 12(1), pp. 35–41.

[36] Parasuraman, R., Mouloua, M., & Molloy, R., 1996, Effects of adaptive task allocation on monitoring of automated systems, *Human factors,* 38(4), pp. 665–679.

[37] Oron-Gilad, T., Ronen, A., & Shinar, D., 2008, Alertness maintaining tasks (AMTs) while driving, *Accident Analysis & Prevention,* 40(3), pp. 851–860.

[38] Schmidt, J., Braunagel, C., Stolzmann, W., & Karrer-Gauss, K., 2016, Driver drowsiness and behavior detection in prolonged conditionally automated drives, In *2016 IEEE Intelligent Vehicles Symposium (IV), Gotenburg, Sweden,* pp. 400–405.

[39] Roth, E., Multer, J., & Raslear, T., 2006, Shared Situation Awareness as a Contributor to High Reliability Performance in Railroad Operations, *Organization Studies,* 27(7), pp. 967–987.

http://ieti.net/TES/
2022, Volume 6, Issue 2, 15-30, DOI: 10.6722/TES.202212_6(2).0003.

ANNEX-1

Table 4. Train driver studies that are included in the analysis.

Nr	Year, First Autor	Independent variables	Subjects	Age (mean or range)	Duration /Distance	Design	Performance measures	Significant Results
1	2005, Lanzilotta [29]	Level of supervisory control	12 students	19-35	3h shifts with training	Within-subject	• Response time to unexpected failures • Accuracy of responses • Situation awareness	• Highest variance in response times to the brake and motor failures with partial automation • The use of partial automation therefore has a negative impact on the consistency of response to unexpected brake and motor failures because the visual attention of the engineer is biased away from the instrument panel. • Supervisory control had no significant effect on the response time to the grade crossing obstruction failures, (attentive to risks occurring at fixed locations)
4	2012, Spring [30]	Level of automation	40 students	22	1h 15 min	Between and within-subject	• Overall vigilance • Vigilance decrement over time	• Infrequent safety critical event detection was poorer at the High LOA compared to the Nil LOA (insignificant for low and intermediate) • High LOA (autopilot) had worse vigilance performance than those who had train control tasks in the lower LOA groups
5.2	2011, Spring [22]	Task demand	40 students	22	1h 15 min	Between and within-subject	• Workload (groups and sessions) • Driving score error • Accurate math quiz responses	• Performance on the Cognitive Vigilance Device secondary task was worse when driving the train, compared to supervising the autopilot • Heart rate measures and SWUC: mental underload was experienced while supervising the train autopilot and operating the Sensory Vigilance Device (SVD). Mental underload was no longer experienced when supervising the train autopilot and operating the CVD. No differences in subjective workload
6	2011, Dunn [14]	Task demand	28 train driver, 28 control	40 driver 32,1 control	3h	Between-subject	• Experience of monotony • Fatigue • Subj. Workload • Speed error • Reaction time	• All groups rated high and low task demands as equally high on scales of monotony, boredom and tiredness, and low in stimulation and engagement • The high demand scenario rated the task significantly higher on the NASA-TLX subscale of mental demands and the task rating scale of effort required

http://ieti.net/TES/
2022, Volume 6, Issue 2, 15-30, DOI: 10.6722/TES.202212_6(2).0003.

								• Participants in the low demand task committed a greater number of errors than those in the high demand task • The performance deteriorates significantly in block 5 for both the low and high demand
7	2014, Large [31]	Task demand	20 control, 7 train drivers	N.A.	10-15min route (2,5h with training)	Between-groups	• Subj. Workload • Overspeed	• No significant difference between drivers and non-drivers in terms of workload and performance • More overspeeds were recorded during the low demand condition compared to the high demand condition
7.2	2014, Large [30]	Task demand (DAS)	20 control, 7 train drivers	N.A.	10-15min route (2,5h with training)	Between-groups	• Subj. Workload	• Using a speed or timetable DAS increased driver workload compared to a control (no DAS) condition for both the low and high demand scenarios.
8	2015, Robinson [9]	Task demand	9 train drivers	51.4	76,3km	Repeated measure	• Reaction time to safety critical events, unexpected AWS irregularities, operational events • Speed choice • Frequency of safety device activations • Subj. workload • Stress arousal (SACL) • Sleepiness (KSS)	• The perceived workload after the mitigation drive was considerably higher than after the baseline drive. • Mean sleepiness ratings increased throughout both drives, but were lower in the mitigation condition both during and after the drive. • There was a main effect of time on task on reported sleepiness. • There was a difference in high arousal scores between both conditions.
9	2016, Stein[20]	Monotony	11 train drivers	39,36	4h	Within-subject	• Reaction time and SPADs (level crossing and stop signals) • Micro-sleep episodes	• The reaction times of in the monotonous condition were higher for the insufficiently secured level crossing task and the stop signals • Sleepiness was higher after the monotonous drive than after the first drive. The KSS ratings after the first half of the monotonous drive are significantly higher than the ratings of KSS after the first drive

http://ieti.net/TES/
2022, Volume 6, Issue 2, 15-30, DOI: 10.6722/TES.202212_6(2).0003.

								• Physiological measures (ECG) showed a decrease of the heart rate (HR) and an increase of the heart rate variability (HRV) after the monotonous ride
10	2016, Brandenburger [28]	Automation	26 train drivers	36,53	105 min (35x3)	Between-subject	• Visual attention (fixations) • Situation awareness	• Number of fixations on DMI higher in ATO condition, but not significant • Number of fixations on DMI higher when the track side view small
11	2017, Brandenburger [32]	Automation	26 train drivers	36,53	120 min (410x3)	Between-subject	• Subjective and objective fatigue	• Clear effect of time across all dependent variables • The heart rate dropped over time and the heart rate variability increased • The KSS scores representing subjectively perceived fatigue also increased significantly over time

Table 5. Key responsibilities of train drivers, associated human factors issues and potential mitigation techniques.

Tasks	Associated issues	Potential effects on the driver	Performance risks	Good practices/potential countermeasures
Incident/obstruction detection	Attention conflict between in-cabin displays and outside view, monotony, time-on-task	Impaired situation awareness (SA level1: perception of relevant information through visual attention) and passive fatigue	Errors of incident detection (track obstructions, collisions at level crossings, staff on tracks)	– Temporal and spatial information on important future points of the route [15] – Onboard assistance systems or obstruction detection technologies at level crossings (does not address underload issues) – System reliability feedback (e.g., assistance systems) [33] – Added cognitive load via secondary tasks during low demand periods (without causing distractions) [9] – Frequent training for cognitive and technical skills [34]

IETI Transactions on Ergonomics and Safety

http://ieti.net/TES/
2022, Volume 6, Issue 2, 15-30, DOI: 10.6722/TES.202212_6(2).0003.

Monitoring system disparities	Low task demand, monotony, time-on-task, complacency	Degraded mental model, passive fatigue, lowered task engagement	Errors in detecting disparities between train behaviour and display information	– System feedback and transparency: Information on important future decision points (braking points, speed limit changes etc.) to verify the correct execution [35] – Promoting anticipatory and proactive driving [13] – System reliability feedback – Commentary driving [23] – Occasional in-cab drills [23] – Frequent training for cognitive and technical skills
Infrequent manual control (e.g., after a system failure)	Under-mobilization of effort after an underloading situation	Lowered task engagement, passive fatigue, de-skilling	Low performance on failure recovery	– Adaptive systems that keep the driver in the control loop instead of as a fallback system [36] – Secondary tasks during low demand periods [37] – Driver performance monitoring [38] – Commentary driving – Frequent simulator training
Operational supervision	Low task demand, complacency	Degraded mental model of the situation	Errors in detecting disparities between train behaviour and operational context. Errors in detecting hazardous situations at station entry/exit	– System feedback and transparency: Information on important future decision points and schedule – Regular and informal communication [39] – Frequent re-training