Predicting Locomotive Crew Performance in Rail Operations with Onboard Human and Automation Assistance

Victoria C. Nneji, *Member, IEEE,* Alexander J. Stimpson, *Member, IEEE,* and M. L. Cummings, *Fellow, IEEE*

***Abstract*—As new technologies are implemented to improve safety and efficiency of rail operations, models are needed to represent the task load of individual crewmembers as well as their interaction with one another to identify periods of extreme workload that could be mitigated through technological interventions. To this end, a systems-theoretic computational model is described to quantitatively model rail engineer and conductor workload to understand the impacts of inserting intelligent automation on different crew configurations and in the locomotive cab. A detailed task analysis served as the basis for identifying tasks performed during transit. Utilizing task characteristics and operating conditions as inputs, a discrete event simulation was designed to predict human operator workload. Results show that, under off-nominal conditions, the presence of automation can impact locomotive engineer performance more than the presence of a freight conductor. However, under nominal conditions, assistance may not be beneficial for human operator performance.**

***Index Terms*— human performance modeling, discrete event simulation, freight rail operations, automation, workload**

I. Introduction

THE U.S. rail transportation industry has a long history of introducing new technologies to meet evolving demands. Innovations such as advanced signaling technology and automated methods for tracking and distributing cars on complex networks of tracks have helped to improve safety by deterring many accidents. They have also facilitated smaller crew sizes and increased workload on crews [1].

Today, human factors are the leading causes of train accidents in the United States [2]. Despite this fact, rules proposed by the Federal Railroad Administration (FRA) in 2016 suggest that safety is significantly improved when the primary operator has another human in the locomotive [3]. Presumably, safety will also be positively impacted when the Congressional mandate set in 2008 for all major railroads to integrate increased levels of automation in the form of positive train control (PTC) is fully enforced [4].

However, previous research in human-automation interaction has demonstrated that additional automation does not necessarily guarantee increased system effectiveness or safety [5], [6]. Often, automating a task within a larger system modifies the task by transferring the operator's workload from one physical or cognitive resource to another, thereby changing the task rather than improving it [7]. Poorly designed automation can contribute to errors and reduce system effectiveness due to implementations that increase cognitive workload.

Cognitive workload is the "level of attentional resources required to meet both objective and subjective performance criteria, which may be mediated by task demands, external support, and past experience" [8]. Workload is of interest in rail operations as the job of human operators in transportation increasingly involves changing demands in cognition and decision making as manual labor is more often allocated to machines [9]. However, while its importance is clear, numerous challenges remain in objectively measuring workload, which have been comprehensively covered [10].

To this end, this research presents a task workload modeling approach, previously applied primarily in aviation settings, applied to freight rail operations. Developed using objective task time data collected during observations of train crews and analytical data collected from interviews with subject matter experts (SMEs), the Simulator of Human Operator Workload (SHOW) represents locomotive crew workload under numerous operating conditions and task loads. SHOW extends previous efforts by developing a computational model to better understand how workload is affected by the introduction of automated technologies, as well as how these advanced technologies could or should inform crew configurations.

II. Literature Review

With increasing automation in the freight rail domain, it is important to understand how operator workload may be affected by introducing new technologies. In support of model development, a literature review provided taxonomies for freight rail operator tasks, uncovered methods to understand and quantitatively describe workload, and provided a method to mathematically express fatigue so that it could be included in a model of workload.

Previous work in the rail domain includes the development of a hierarchical task network to abstractly represent workload [11]. To structure that network, researchers synthesized cognitive task analyses of engineer [12], conductor [13], and dispatcher [14] roles and summarized the results in concept maps. They then verified the accuracy of the maps through

Manuscript submitted October 2017. This work was supported by the U.S. Department of Transportation under Grant DTFR53115C00017.

V.C. Nneji, A. Stimpson, and M. L. Cummings are with the Humans & Autonomy Lab of Duke Robotics, Duke University, Durham, NC 27708 USA (e-mails: victoria.nneji@duke.edu, alexander.stimpson@duke.edu, m.cummings@duke.edu).

interviews with subject matter experts (SMEs) and used the concept maps to develop an abstraction hierarchy, as proposed in earlier work [15], representing specific tasks and generalized functions executed by engineers and conductors for safe and effective rail transport.

The hierarchical task network provided valuable insight into the causes of operator workload. However, given the insertion of new technologies into rail operations, an additional method is needed to build a model of workload that allows for concept evaluation across different operating conditions, especially as such changes could significantly impact operator workload and system performance. Mathematical models represent an objective-analytical approach to consistently calculate human operator workload across different conditions and time scales [16] that are applicable to the railroad industry. Work in human supervisory control [17], human-machine interaction [18], air traffic control [19] have suggested time-on-task as a reliable estimation of workload.

Time-on-task is defined as the duration an operator's attentional resources are actively used to meet functional requirements. Wierwille and Eggemeier [10] suggest that time-on-task estimations can reveal areas of heightened workload that may be missed in other forms of assessment. Cummings and Guerlain [20] and Donmez, Nehme, and Cummings [21] extended the use of task time data by incorporating a utilization metric as a means of objectively measuring high as well as low workload. Their utilization metric is defined as percentage of time an operator spends on task performance out of total operation time. Maximum utilization is 100%, at which point there no additional capacity available for a human operator to allocate toward accomplishing tasks. When utilization levels are too high, operators may be too busy to accumulate the information required to maintain situation awareness (SA). Similarly, when operators are underutilized, they could overlook information from the environment due to low engagement, also leading to poor SA. Levels of utilization below 30% have been associated with poor performance due to boredom and distraction while a 70% utilization threshold has been used to indicate the upper bound of optimal workload [22]–[25]. For over a century, human factors researchers have used the Yerkes-Dodson inverted-U theory to conceptualize the relationship of workload levels to performance where performance declines during extreme workload periods including under- or over-utilization [21]. The goal for rail operations is to optimally man crews such that each operator is moderately utilized.

Even under optimal staffing, operator performance may dwindle over the course of a shift. Fatigue is a major consideration in rail operations. While fatigue has been shown to deplete the level of attentional resource available for operators attending to tasks [26], it can also be alleviated when operators consistently recuperate capacity by getting eight consecutive hours of sleep between work periods [27]. Part 228 of the Code of Federal Regulations 49 [28] dictates that train employees must not work more than 12 consecutive hours without 10 consecutive hours off duty or to have less than 8 consecutive hours off duty in a 24-hour period. However, the current state of the industry is such that many operators in freight rail services have unpredictable work schedules, which may lead to repeatedly disrupted sleep cycles and rest periods below eight hours [29]. In turn, fatigue can lead freight rail operators into dangerous states prone to human error from delaying, neglecting, or incorrectly performing time-sensitive tasks [30].

With significant research efforts into the effects of fatigue on operator performance [31], exactly how to use such research to inform quantitative performance models requires additional investigation. Hursh, Redmond, Johnson, Thorne et al. [32] present a mathematical equation which computes cognitive performance capacity decay over consecutive hours of wakefulness. It approximates to a linear function that projects 1% increases in time-on-task with each hour of a shift when transformed inversely.

To model this effect of fatigue and other effects on rail engineer and conductor workload levels, a model is needed that represents both the nature and variety of tasks expected of both. Because of the temporal nature of shift work experienced by rail crews, we chose to develop a workload model through discrete event simulation (DES). DES has been used for years in modeling manufacturing and health care services. It allows designers to realize areas that cause undesired delays in their systems. DES has been used in some studies of railway operations [33]–[35], however, the industry needs an approach to model human workload across the rail system, including engineers, conductors, and dispatchers, with a special emphasis on the role of increasing automation. With pending changes in operational practices, (e.g., increased automation and changing crew size), stakeholders need a tool for rapid exploration of various crew workload and tasking models. Important future planning questions include: what tasks could potentially be offloaded from the locomotive engineers onto other agents such as PTC or remote dispatch, and what would be the workload implications for human operators?

This paper describes the design and analysis of a functional DES of freight rail operations to model the workload of human operators. Using DES is advantageous for this setting in that it allows us to have a time-based representation of human processes at a task-level. Therefore, we began this modeling effort by collecting time-on-task data of real freight rail operations.

III. Empirical Data Collection

Cognitive task analyses were conducted through direct observation, and structured and unstructured interviews with SMEs to identify and categorize in-cab crew tasks. Using present-day operations as a baseline, a typical in-cab crew can include a team of two or single-person operations, with the engineer defined here as the primary operator. Based on a framework proposed by Subrahmaniyan et al. [11] that generalized functions to represent work performed by the engineer and conductor, a core set of tasks were defined as shown in Table I.

The work of the engineer is more than just maneuvering the locomotive control system. It also requires paying attention to the environment, in-cab displays, and radio communications with dispatch to maintain situation awareness and plan. On the other hand, the work of the conductor is to supervise the train conditions on the ground at terminal points and remain attentive to the engineer while the train is in motion in the case of an emergency, when action could be needed. When the conductor engages in planning, monitoring inside, or exception handling tasks, they function as a backup since tasks of those type are primarily allocated to the engineer who has been trained to handle them as a single operator.

TABLE I. CORE TASKS OF FREIGHT RAIL CREW

Task Type	Description
Communicating	Filtering through relevant information for the operation and communicating information that may impact the macro-level network of operations
Exception Handling	Attending to unexpected or unusual situations that must be handled to continue with the trip
Paperwork	Reviewing and recording operating conditions
Maintenance-of-Way Interactions	Maintaining situational awareness of other crews along the track
Temporary Speed Restrictions	Recalling information issued on track bulletins and adapting to updates while train is in motion
Signal Response Management	Maintaining attentiveness to direction from track signaling system and responsive to proper control system within a safely allotted time
Monitoring Inside	Maintaining attentiveness to informational displays and the engineer's performance to maintain a safe operation
Monitoring Outside	Maintaining attentiveness to warnings and environmental conditions that may affect operations
Planning	Maneuvering locomotive control system for throttle, braking and other subtasks like horn-blowing before railroad crossing

We directly observed locomotive engineers during two 2.5-hour ride-alongs at a private Canadian regional railroad that operates with single-person crews. During the observations, our team noted a description and time for each task observed (e.g., blowing horn, 00:15:11-00:15:16). In total, approximately 50 tasks were logged per hour.

The ride-alongs were followed by structured interviews with five SMEs in the railroad's operations, training, dispatch, and regulations along with two labor union representatives from a Canadian Rail United Transportation Union chapter with experience across all personnel ranging from 3 to 20 years. Video recordings from the onboard observations were reviewed to identify times spent on each of the nine high-level tasks performed by the operators throughout different phases of the shift (e.g., communicating, paperwork).

Interviews following observations provided clarification of tasks observed during normal railroad operations. At the Amtrak Training Center in Wilmington, Delaware, additional structured interviews were conducted with four engineer and conductor SMEs along with observations of engineer task performance in three physical cab simulators and one trainee class to supplement our dataset.

One important result from these observations was the elucidation of different phases of operation. A phase is defined by a change in operational behavior which drives either different tasks for an operator, or substantially different frequencies of tasks. From the cognitive task analyses and observations, we identified three phases for freight rail operators: startup, full motion, and yard.

The startup phase includes the first 30 minutes of an engineer's shift in the train. Startup includes tasks supporting regulatory requirements, such as communicating with dispatch and testing the emergency braking system so, just as in the real world, there is a chance that a task comes in last minute that takes longer than expected to complete.

During full motion the engineer drives the locomotive to its destination and this typically is the longest phase of the trip. There is a marker at 30 minutes before the end of the engineer's shift to represent the transition from full motion to the yard phase, which involves some change in how tasks behave in the system. This final phase is also the source of the highest rates of accidents [36].

IV. DESCRIPTIVE MODEL DEVELOPMENT

The results of these cognitive task analyses of the engineer and conductor functions were used to construct a descriptive temporally-based task model of crew responsibilities, including frequencies and durations for the observed, and estimates for the unobserved, engineer and conductor tasks.

The results from this descriptive model were validated by subject matter experts for four defined scenarios that described workload of an engineer working alone or with a conductor during a 5-hour shift under nominal or off-nominal conditions under both crew complements. One scenario is presented in Fig. 1.

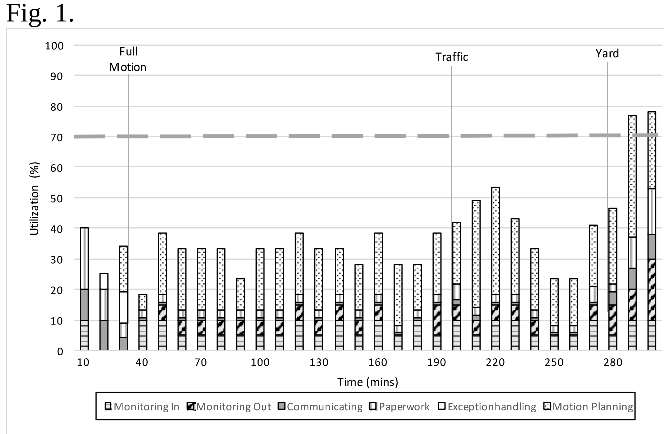

Fig. 1. Percent of engineer utilization in a one-person crew in 10-minute intervals over a 5-hour shift during the nominal condition.

In Fig. 1, the scenario of an engineer operating alone during nominal conditions is presented. Task times are described over 10-minute time ranges. The three markers labeled *Full Motion* beginning in the first 30 minutes, *Traffic* beginning at minute 200, and *Yard* approaching the final 30 minutes represent changes in tasks and frequencies of tasks. For example, the *Traffic* event represents the train approaching other vehicles on the track.

The engineer's tasks are summarized per the categories in Table I: (1) monitoring in, (2) monitoring out, (3) communicating, (4) paperwork, (5) exception handling, and (6)

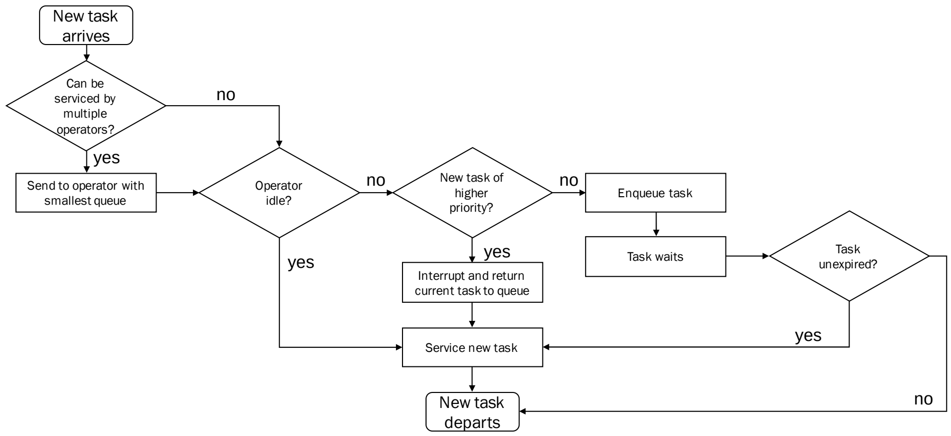

Fig. 2. SHOW process flow from task arrival to departure from freight rail operator system.

motion planning. Motion planning consists of maintenance-of-way, temporary speed restrictions, signal response management, and planning tasks. The model shows that almost half, 46%, of the engineer's workload can be attributed to *motion planning*. Motion planning involves manipulating the locomotive control system and preparing in advance for mechanisms required to move the train towards a destination. Monitoring the displays within the cab makes *monitoring inside* an important secondary task that accounts for 22% of their total time on task during the 5-hour shift.

In this scenario that represents nominal operations over a 5-hour period, typical of short haul lines, the engineer is responsible for fulfilling all onboard tasks, and Fig. 1 shows that utilization ranges between 18% and 78%, with an average of 37% utilization. In this scenario, the engineer spends 20 minutes above 70% and 80 minutes below the 30% utilization threshold. Using these and similar descriptive results from the remaining scenarios, a predictive DES model was developed to simulate numerous operating conditions including the variability that can occur in present and future freight rail system configurations.

V. Simulator of Human Operator Workload

The resulting DES model is called the Simulator of Human Operator Workload (SHOW), which models human operators and assistive agents as serial processors of complex tasks and records the basic units of key performance metrics.

The simulation begins when the first task arrives in the system (based on a stochastic arrival distribution) and then is assigned to an operator's queue based on the operator's functional capability to handle that type of task. Operators include humans (i.e. conductor and/or engineer) and automated agents (i.e. PTC and/or cruise control technologies). Tasks can queue, awaiting operator availability for servicing. In the process of service, the task may be interrupted by another task of higher priority and thus returned to wait in queue due to preemptive priority scheduling. Finally, the task departs the system when it is completely serviced or is expired.

The process of a task flowing through the freight rail operational system is illustrated in Fig. 2. At any point in this process, the task may expire before service, at which point it departs the system. A task may also drop if the phase changes or the trip ends while it is waiting in a queue. For example, at 30 minutes, no additional startup tasks can enter the queue and engineers must complete any tasks already in the queue before they can begin full motion phase.

The simulation allows tasks to have different arrival rates and expiration times by phase as appropriate. This is intended to mimic real-world operations, where types of tasks may differ between phases. For example, a maintenance of way interaction task would only arrive in the latter two phases of a trip.

The sources of system stochasticity are task inter-arrival and operator service times which are random observations generated from probability distribution functions. Each replication in a simulation run pulls from the random distributions, which simulates the variability of event times. More simulated trips provide a more robust set of results although the simulation's processing time will increase [37]. The simulation tracks several different statistics as shown in Table II, computed for each human operator across replications.

Utilization, the principal measure, is used as a proxy for workload. The simulation records the utilization for each operator throughout the shift and results are reported in 10 minute intervals. This output can be used to validate the simulation model output data when compared to the empirical data structure. The expired/completed tasks and average wait time outputs provide additional statistics to infer system performance metrics of effectiveness and efficiency.

TABLE II. Simulation Output Statistics

Output Statistic	Description	Purpose
Utilization	Time on task divided by total (10 min) time interval	Proxy for workload
Expired/Completed Tasks	Tasks that were not versus tasks that were completed by the operator	Identify which tasks are more likely to be missed; throughput
Average Wait Time	Average time each task waited in the operator's queue	Identify which tasks spend the most time waiting

VI. Simulation Validation

The validation process establishes the model as a reasonable reflection of the real system it attempts to simulate [38].

Therefore, once the stand-alone simulation software SHOW, developed in C++, was verified with Rockwell Automation's off-the-shelf Arena version 14.7 package, a three-step process was followed:

1. Statistical comparison with descriptive model for external results validation
2. Subject matter expert review for face validation
3. Sensitivity analysis for internal validation

Each of these steps are intended to build confidence in the model's real-world usefulness [39]. The first step proved that the model replicated results found in a present-day system observed and described from data collected in the field. To validate the output from the overall simulation model, a Kolmogorov-Smirnov statistical-goodness-of-fit test was performed to determine if its data closely compared to that of the actual system described. The test failed to reject the null hypothesis (α=.05) that the samples from our descriptive and predictive models are drawn from different distributions during the *Full Motion* phase of the shift. D=.2917 (< D_{α}), P = .216.

The second step involved gathering qualitative feedback from SMEs with experience in freight rail operations. SMEs included seven freight rail original equipment manufacturers, researchers, and operators within the locomotive as well as in a managerial level—both current and retired. The SME review supported the model's results that conductors have overall lower workload than engineers. SMEs also attested to operators experience higher workload on average during the final phase of each trip. In general, those with careers directly involving freight rail operations found the results of both locomotive engineer and freight conductor to not be significantly different from their own experiences. However, some researchers who had experience using less quantitative methods of studying workload expressed skepticism about the utilization levels reported to be lower than they imagined. Discussion with the groups revealed that the model may be more valid if shown to be applicable to other settings in freight rail operations such as dispatch centers.

Finally, a sensitivity analysis was performed to discover which model factors may significantly impact key performance indicators (KPIs). The goal of the analysis was to investigate how levels of variation in parameters impact the deviation of average utilization, wait time, throughput rate, and expiration rate output measurements from the baseline results.

Overall, statistical sensitivity analysis revealed that variations in inter-arrival time significantly change mean results for utilization and throughput rate. Variations in service time significantly change mean results for utilization. And only extreme variations in traffic parameters significantly change mean utilization and throughput rate, therefore this validates conditions to test operations for off-nominal scenarios. In contrast, none of the variations tested in the fatigue parameter significantly changes any KPIs.

As with any model, there are several limitations based on assumptions. First, the task distributions and parameters may not accurately reflect all freight rail scenarios due to the limited scope of the ride-alongs used for parameter estimation. Second, the model does not yet incorporate the possibility for error within the processing of tasks. This idea is in development for future work, since such an addition would more accurately reflect the real world. Third, the model does not account for some characteristics of human operators that could impact performance such as the hours of sleep prior to the shift. Finally, the simulation models a single train. However, the validation strategy demonstrated that SHOW, a novel approach for modeling operator workload in freight rail environments, is credible for simulating current operations and studying potential future system changes.

VII. Simulation Analysis

The validated model was used to conduct a comparative analysis of operator workload in futuristic system reconfigurations. Our objective is to identify which combinations of human crews and automated assistants yield better performance for the system.

Four levels of assistance and two levels of conditions were defined (Table III). When the engineer has no assistance, the engineer is responsible for all tasks alone. Human assistance refers to an onboard freight conductor who can offload signal response management, monitoring inside, and planning tasks from the engineer. Automation refers to positive train control (PTC) and cruise control [40] technologies which can offload emergency braking, signal response management, monitoring inside, and planning ahead tasks from the engineer. The level of *all* assistance means that the engineer is supported both by the conductor and automation. For any shared tasks, the engineer is the primary operator but when busy, the task is allocated to the conductor, and then to automation.

A penalty function (1) was defined to identify how the locomotive engineer's performance may be influenced by levels of automated and human assistants under nominal and off-nominal conditions. The penalty function was applied across 1000 replications to analyze the distribution of the results for each of the eight experiments presented in Table III. The penalty function approximates the Yerkes-Dodson Law (Fig. 3) which establishes performance as an inverted parabolic function of workload. In this case, penalty is a function of utilization.

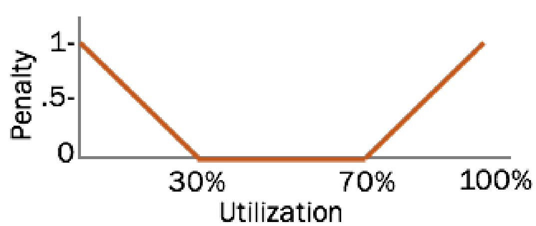

Fig. 3. The penalty function of operator utilization.

The thresholds for potential decrements in performance are marked at 30% and 70% utilization based on results from previous experiments [22]–[25]. At those points, a penalty value (0,1] is computed along a linear slope of -1 for low workload and +1 for high workload. For each 10-minute interval recorded in the simulation, the following algorithm generates a penalty value:

$$
\begin{aligned}
&\mathit{If\ Utilization} < 30\%, && (1)\\
&\qquad \mathit{Penalty} = 1 - 3.33 * \mathit{Utilization}\\
&\mathit{Else\ if\ Utilization} > 70\%,\\
&\qquad \mathit{Penalty} = 3.33 * \mathit{Utilization} - 2.33\\
&\quad \mathit{Else}, \mathit{Penalty} = 0.
\end{aligned}
$$

TABLE III. DESIGN OF EXPERIMENTS

Experiment	Assistance	Condition
	Levels	
1	None	Nominal
2	All	Nominal
3	Automation	Nominal
4	Conductor	Nominal
5	None	Off-nominal
6	All	Off-nominal
7	Automation	Off-nominal
8	Conductor	Off-nominal

Nominal conditions represent *normal* operations as observed and validated. Off-nominal conditions represent a *worst* day an operator may have, with respect to traffic events, leading to overall high task load from 3.5 times the nominal task arrival rates.

A one-way analysis of variance (α=.001) was used to determine whether penalty values from extremely high workload associated with the off-nominal conditions differed when the locomotive engineer was alone, supported by all assistance, supported by automation only, or supported by conductor. Fig. 4 shows that the engineer assisted with a combination of automation and human support exhibited the best performance with penalty (M=2.69, S.D.=1.63), somewhat lower performance with automated assistance only (M=2.71, S.D.=1.61), poorer performance with human assistance only (M=3.21, S.D.=2.18), and worst performance without any assistance (M=3.55, S.D.=1.95).

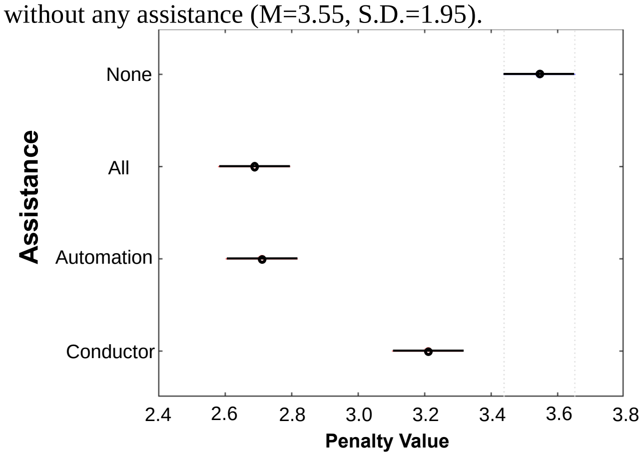

Fig. 4. Mean penalty values with comparison intervals (α=.001) for high workload experienced by engineer in off-nominal condition with each level of assistance. "None" level of assistance was significantly different from every other level.

The analysis showed significant differences among the groups of data. F(3,3996) = 49.9, p < 0.00001. Post-hoc Tukey's honest significant difference tests showed that engineer performance during no-assistance operations differed significantly from each of the other three settings, but the difference between all- versus automated-assistance was not statistically significant. So, while on average all assistance leads to better performance, there is some variation in the results. The simulation results show that the conductor combined with automation does not make much of a difference versus automation assistance alone (p > .001), although the presence of automation with or without the conductor made much more of a difference than with the conductor alone.

However, under nominal conditions any form of assistance was found to lead to significantly higher penalty values due to extremely low workload experienced by the locomotive engineer as shown in Fig. 5. A one-way analysis of variance was used to determine whether penalty values differed when the locomotive engineer was alone, supported by all assistance, supported by automation only, or supported by human only.

The analysis showed significant differences among the groups of data. F(3,3996) = 75.77, p < 0.00001. A lone engineer exhibited the best performance based on the penalty score (M=10.25, S.D.=1.48), somewhat lower performance with human assistance (M=10.96, S.D.=1.48), poorer performance with automation (M=11.06, S.D.=1.41), and worst performance when assisted with both human and automation (M=11.09, S.D.=1.41). Post-hoc Tukey's honest significant difference tests showed that engineer performance during no-assistance operations differed significantly from each of the other three configurations, but the difference between all- versus automation versus human-assistance was not statistically significant. So, under nominal conditions and low task load, engineers working alone performed better than with any assistants, while under off-nominal conditions engineers working with automation (with or without human assistance) had the best performance of the tested conditions.

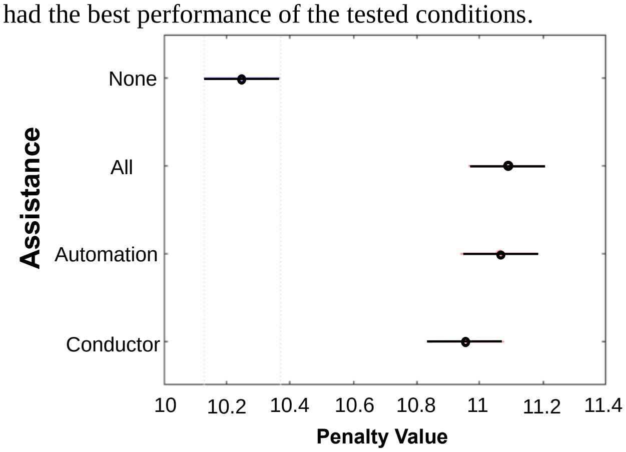

Fig. 5. Mean penalty values with comparison intervals (α=.001) for low workload experienced by engineer in nominal conditions with each level of assistance. "None" level of assistance was significantly different from every other level.

Interestingly, the average penalty acquired under nominal conditions (M=10.84) was significantly higher than under off-nominal conditions (M = 3.04). In referring to the penalty function (1), this may be attributed to extended periods of time at or below 30% utilization during the shift.

VIII. Discussion & Conclusion

The Simulator of Human Operator Workload (SHOW) was developed as a discrete event simulation to model the impact of novel technologies on the locomotive cab environment. This process involved gathering empirical data from key stakeholders in the railroad industry. SHOW was validated through statistical goodness-of-fit tests, subject matter expert review, and sensitivity analysis.

In comparing performance across alternative system configurations in SHOW, results suggest that the impact of automation on performance is more helpful than a freight conductor at moderating locomotive engineer workload during off-nominal conditions. However, that same automation could be detrimental to operator performance during nominal conditions.

At this crucial moment in the industry with major investments already in installing automation and making decisions about the future workforce [41], a simulation model like SHOW is needed to investigate the potential implications of human- and automated- assistance on the primary operator's workload across operating conditions that include novel automation technology. Particularly in the setting of freight rail operations, as levels of automation increase, we can expect boredom and the associated reductions in performance to also increase [23][42]. PTC, combined with other technologies like energy management cruise control systems, would provide efficiencies that lead many railroads to believe that the second crewmember is unnecessary [43]. The results from SHOW demonstrate that we cannot take a blanket approach to improving safety conditions. Consideration is required to identify when and how automation could be of benefit without detriment to human performance across the spectrum of operational conditions.

Future research should be devoted to better understanding the unexpected challenges that automation and human assistance may bring to a primary operator's performance. Given the debate as to what constitutes optimal staffing on a train, another consideration is shifting the engineer off-board to control multiple trains on a supervisory level. When thoughtfully designed, human supervisory control has been shown to be beneficial to system performance in similar operational settings with adaptive automation [44].

Our future focus will be on that of a railroad dispatch center, to explore how new reallocation schemes between remote human operators and artificially intelligent agents may support overall system performance across a network of trains. This would allow an analyst to see how small, local changes could impact the broader, interconnected system.

Acknowledgment

The views and opinions expressed are those of the authors and do not necessarily reflect the views of the US DOT. Branch Vincent was a key member of our team during the development phase and Dr. Michael Clamann helped in the data collection phase. Finally, we thank the subject matter experts who shared time and wisdom with us.

References

[1] C. D. Martland, “Workload Measurement and Train Crew Consist Adjustments: The Boston & Maine Experience,” *Transp. J.*, vol. 21, no. 3, pp. 34–57, 1982.

[2] FRA, “Summary of Accident/Incident Counts from January-December 2015 Priliminary Data,” *U.S. Department of Transportation Research*, 2016. [Online]. Available: http://safetydata.fra.dot.gov/officeofsafety/publicsite/summary.aspx . [Accessed: 01-Jan-2016].

[3] FRA, *Notice of Proposed Rule Making: Train Crew Staffing*. 2016, pp. 1–205.

[4] 110th US Congress, *Rail Safety Improvement Act (RSIA)*. 2008, p. Public Law 110-432.

[5] R. Parasuraman and V. Riley, “Humans and Automation: Use, Misuse, Disuse, Abuse,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 39, no. 2, pp. 230–253, 1997.

[6] M. L. Cummings and J. Ryan, “Shared Authority Concerns in Automated Driving Applications,” *J. Ergon.*, vol. 3, pp. 1–4, 2013.

[7] A. Sebok, C. D. Wickens, and L. Laux, “Human-Automation Interaction : Lessons Learned from the Aviation Industry and Recommendations to the Rail Industry for Regulatory Guidance,” Washington, D.C., 2015.

[8] M. S. Young and N. A. Stanton, “Mental Workload,” in *Handbook of Human Factors and Ergonomics Methods*, CRC Press LLC, 2005, pp. 391–399.

[9] W. T. Singleton, *The mind at work: Psychological ergonomics.* 1989.

[10] W. W. Wierwille and F. T. Eggemeier, “Recommendations for Mental Workload Measurement in a Test and Evaluation Environment,” *Hum. Factors*, vol. 35, no. 2, pp. 263–281, 1993.

[11] N. Subrahmaniyan, A. Liu, B. Miller, H. Groshong, and J. D. Brooks, “Investigation of New Roles for Humans and Automation in Rail : An Interim Report,” Washington, D.C., 2014.

[12] E. Roth and J. Multer, “Technology Implications of a Cognitive Task Analysis for Locomotive Engineers,” Washington, D.C., 2007.

[13] H. Rosenhand, E. Roth, and J. Multer, “Cognitive and Collaborative Demands of Freight Conductor Activities Results and Implications of a Cognitive Task Analysis,” Washington, D.C., 2012.

[14] E. M. Roth, N. Malsch, and J. Multer, “Understanding How Train Dispatchers Manage and Control Trains: Results of a Cognitive Task Analysis,” Washington, D.C., 2001.

[15] K. J. Vicente, “Cognitive Work Analysis,” *Analysis*, vol. 17, no. 3, pp. 313–21, 1999.

[16] C. Rusnock, B. Borghetti, and I. McQuaid, “Objective-Analytical Measures of Workload – the Third Pillar of Workload Triangulation?,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9183, pp. 124–135, 2015.

[17] M. L. Cummings and C. E. Nehme, “Modeling the impact of workload in network centric supervisory control settings,” *Neurocognitive Physiol. factors Dur. high-tempo Oper.*, pp. 23–40, 2010.

[18] W. B. Rouse, “Models of human problem solving: Detection, diagnosis, and compensation for system failures,” *Automatica*, vol. 19, no. 6, pp. 613–625, 1983.

[19] D. K. Schmidt, “A queuing analysis of the air traffic controller’s work load,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 8, no. 6. pp. 492–498, 1978.

[20] M. L. Cummings and S. Guerlain, “Developing operator capacity estimates for supervisory control of autonomous vehicles.,” *Hum. Factors*, vol. 49, no. 1, pp. 1–15, 2007.

[21] B. Donmez, C. Nehme, and M. L. Cummings, “Modeling Workload Impact in Multiple Unmanned Vehicle Supervisory Control,” *IEEE Trans. Syst. Man. Cybern.*, vol. 40, no. 6, pp. 1180–1190, 2010.

[22] R. M. Yerkes and J. D. Dodson, “The Relation of Strength of Stimulus to Rapidity of Habit-Formation,” *J. Comp. Neurol. Psychol.*, vol. 18, no. 5, pp. 459–482, 1908.

[23] M. L. Cummings, F. Gao, and K. M. Thornburg, “Boredom in the workplace: A new look at an old problem.,” *Hum. Factors*, vol. 58, no. 2, pp. 279–300, 2016.

[24] M. L. Cummings, C. Mastracchio, K. M. Thornburg, and A. Mkrtchyan, “Boredom and distraction in multiple unmanned vehicle supervisory control,” *Interact. Comput.*, vol. 25, no. 1, pp. 34–47, 2013.

[25] W. B. Rouse, *Systems engineering models of human-machine interaction*, vol. 6. North-Holland, 1980.

[26] G. D. Roach, A. Fletcher, and D. Dawson, “A Model to Predict work-related fatigue based on hours of work,” *Aviat. Sp. Environ. Med.*, vol. 75, no. SUPPL.1, pp. 61–69, 2004.

[27] P. Sherry, G. Belenky, and S. Folkard, “Hours of Service Regulations in the U.S. Railroad Industry: Time for a Change,” *Transp. Law J.*, vol. 33, no. 3, pp. 295–338, 2006.

[28] Federal Railroad Administration, *Appendix A to Part 228--Requirements of the Hours of Service Act: Statement of Agency Policy and Interpretation*. 2012, pp. 463–467.

[29] B. Zimmerman, “Railway Crew Rostering and its Effects on the Planned Workforce: Behvioural Operations in the European Railway Industry,” Erasmus University Rotterdam, 2012.

[30] J. Dorrian, G. D. Roach, A. Fletcher, and D. Dawson, “Simulated train driving: Fatigue, self-awareness and cognitive disengagement,” *Appl. Ergon.*, vol. 38, no. 2, pp. 155–166, 2007.

[31] P. A. Desmond, M. C. Neubauer, G. Matthews, and P. A. Hancock, *The Handbook of Operator Fatigue*. Ashgate Publishing, Ltd., 2012.

[32] S. R. Hursh, D. P. Redmond, M. L. Johnson, D. R. Thorne, G. Belenky, T. J. Balkin, W. F. Storm, J. C. Miller, and D. R. Eddy, “Fatigue models for applied research in warfighting,” *Aviat. Space. Environ. Med.*, vol. 75, no. 3, Suppl., pp. A44-53, 2004.

[33] M. H. Cha and D. Mun, “Discrete event simulation of Maglev transport considering traffic waves,” *J. Comput. Des. Eng.*, vol. 1, no. 4, pp. 233–242, 2014.

[34] C. Woroniuk and M. Marinov, “Simulation modelling to analyse the current level of utilisation of sections along a rail route,” *J. Transp. Lit.*, vol. 7, no. 2, pp. 235–252, 2013.

[35] A. Nash and D. Huerlimann, “Railroad simulation using OpenTrack,” *Comput. Railw. IX*, pp. 45–54, 2004.

[36] FRA, “Federal Railroad Administration Status Update on Positive Train Control Implementation,” Washington, D.C., 2016.

[37] S. Robinson, *Simulation: the practice of model development and use*. 2004.

[38] A. M. Law and W. D. Kelton, “Simulation Modeling and Analysis,” *ACM Transactions on Modeling and Computer Simulation*, vol. 2, no. 2. pp. 1370–1375, 2000.

[39] F. Leal, R. F. D. S. Costa, J. A. B. Montevechi, D. A. De Almeida, and F. A. S. Marins, “A practical guide for operational validation of discrete simulation models,” *Pesqui. Operacional*, vol. 31, no. 1, pp. 57–77, 2011.

[40] C. R. Cole, S. Simson, M. McClanachan, T. McLeod, and J. Phelan, “Train cruise control,” in *In proceeding of the 9th International Heavy Haul Conference*, 2009, pp. 638–645.

[41] G. Peters, “Are driverless freight trains safe?,” *Railway Technology*, 20-Sep-2016.

[42] D. Sussman and M. Coplen, “Fatigue and Alertness in the United States Railroad Industry Part I: The Nature of the Problem,” *Transp. Res. part F traffic Psychol. Behav.*, vol. F, no. 48, pp. 1–8, 2001.

[43] B. Stephens, “CSX’s Ward says one-person crews are ‘inevitable,’” *Trains Magazine*, 2017. [Online]. Available: http://trn.trains.com/news/news-wire/2017/01/18-csx-ward-one-person-crew. [Accessed: 21-Oct-2017].

[44] C. A. Miller and R. Parasuraman, “Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 49, no. 1, pp. 57–75, 2007.

Victoria Chibuogu Nneji (M’16) is a creative thinker and thoughtful creator apt to find paths in complexity toward solutions in service to her communities. She was born in Lagos, Nigeria and raised in Durham, North Carolina, USA where she is currently a Ph.D. student in mechanical engineering-robotics at Duke University. She graduated with a Master of Engineering Management, also at Duke, in 2015, and a B.S. in Applied Mathematics at Columbia University in New York City. As a research assistant in the Humans & Autonomy Lab at Duke she is interested in how we can better design advanced mobility and logistics systems with humans in mind. Her work spans trains, planes, and automobiles.

Alexander Stimpson (M’14) received his B.S. degree in biological engineering from the University of Florida, Gainesville, FL, USA, in 2007, and the S.M. degree in Aeronautics and Astronautics from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2011. His dissertation work focused on the application of machine learning models to inform training assessment and intervention. His current research interests include human supervisory control, decision support systems, artificial intelligence, and data mining. He is currently a postdoctoral researcher for the Humans and Autonomy Laboratory at Duke University.

Mary (Missy) Cummings (SM’03) received her Ph.D. in Systems Engineering from the University of Virginia in 2004. She is currently an associate professor in the Duke University Department of Mechanical Engineering and Materials Science, the Duke Institute of Brain Sciences, and the Duke Electrical and Computer Engineering Department. She is the director of the Duke Humans and Autonomy Laboratory.