Applied Ergonomics 114 (2024) 104154

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

Modeling novice law enforcement officers' interaction with in-vehicle technology

Junho Park, David Wozniak, Maryam Zahabi *

Industrial and Systems Engineering, Texas A&M University, College Station, TX, USA

A R T I C L E I N F O

Keywords:
Novice
Cognitive modeling
Law enforcement
Human performance modeling

A B S T R A C T

Cognitive performance models have been used in several human factors domains such as driving and human-computer interaction. However, most models are limited to expert performance with rough adjustments to consider novices despite prior studies suggesting novices' cognitive, perceptual, and motor behaviors are different from experts. The objective of this study was to develop a cognitive performance model for novice law enforcement officers (N-CPM) to model their performance and memory load while interacting with in-vehicle technology. The model was validated based on a ride-along study with 10 novice law enforcement officers (nLEOs). The findings suggested that there were no significant differences between the N-CPM and observation data in most cases, while the results of the benchmark model were different from that of N-CPM. The model can be applied to improve future nLEO's patrol mission performance through redesigning in-vehicle technologies and training methods to reduce their workload and driving distraction.

1. Introduction

Motor vehicle crashes (MVCs) are a major cause of death in the U.S. In 2021, an estimated 46,000 people lost their lives to car crashes and about 5.2 million people were seriously injured in crashes (NSC, 2022). MVCs are also a leading cause of line-of-duty deaths for public safety workers, especially law enforcement officers (LEOs) (BLS, 2020). LEOs are involved in a significantly higher number of fatal MVCs as compared to firefighters and emergency medical services workers (BLS, 2019). The rate of LEO MVCs is also 2.5 times higher than the national average among all occupations (Maguire et al., 2002). The main contributors to these crashes include officers' use of in-vehicle technologies while driving (Yager et al., 2015), fatigue (Vila and Kenney, 2002), and lack of sufficient training in handling high-demand situations (e.g., pursuit situations, multi-tasking) (Hembroff et al., 2018). Prior investigations (Park et al., 2020; Shupsky et al., 2021; Zahabi and Kaber, 2018a, 2018b; Zahabi et al., 2020a) have identified the mobile computer terminal (MCT) (a laptop that provides real-time information to LEOs) and radio as the most important and frequently used in-vehicle technologies while driving. Use of these technologies has increased LEOs' distraction and cognitive load while driving (Shahini et al., 2020).

1.1. Modeling novice law enforcement officers' (nLEOs) performance

Cognitive performance modeling is an approach to model human information processing or a pattern of actions carried out to satisfy an objective (Zhang and Wu, 2017). Creating a cognitive performance model (CPM) provides several advantages over human subject experiments, particularly during the initial phases of a design process (Park and Zahabi, 2022). CPM is a faster and safer approach compared to experimental studies because it minimizes human subjects' involvement. Furthermore, the models can quantify and predict human behaviors in natural tasks while also being easily modifiable (Salvucci et al., 2005; Zahabi et al., 2019b; Zhang and Wu, 2017). The CPM approach has been applied in a diverse range of human factors domains, including aerospace systems (Redding, 1992), augmented cognition (Fincham, 2005), computer systems (St. Amant et al., 2004), healthcare (Zahabi and Lyman, 2019), human-AI-robot teaming (Dudzik, 2019), perception and performance (Jeffrey Bolkhovsky et al., 2018), surface transportation (Tsimhoni and Reed, 2007), and user testing and evaluation (Oyewole et al., 2011). For surface transportation, using the CPM method is a safer approach than naturalistic studies (due to safety concerns) or driving simulation experiments which might cause simulator sickness.

* Corresponding author. Industrial and Systems Engineering Department, Texas A&M University, Emerging Technologies Building, College Station, TX, 77843-3131, USA.

E-mail address: mzahabi@tamu.edu (M. Zahabi).

https://doi.org/10.1016/j.apergo.2023.104154
Received 2 June 2023; Received in revised form 18 September 2023; Accepted 12 October 2023
Available online 24 October 2023
0003-6870/

Considering the seriousness of LEO MVCs and the main contributors of these crashes, there are several reasons why modeling the behavior of novice LEOs (nLEOs) is important. First, previous models cannot model time-sensitive situations such as emergency operations of nLEOs because they assume experts perform all tasks in a specific order and do not make mistakes. The new model for novices should reflect their cognitive, perceptual, and motor demands while driving. Additionally, conventional training approaches used to instruct nLEOs such as training videos, classroom education, vehicle owner manuals, and on-the-road skill training have been demonstrably ineffective at improving multi-tasking performance (Christie, 2001; Peck, 2011). The high cognitive workload created by patrol situations reduces the effectiveness of recall for information used for performing secondary tasks such as managing information on a MCT or responding to calls on a radio (Galy et al., 2012). Models of nLEOs' performance in these demanding situations can estimate their recall probability and ultimately, enhance the design of training methods and in-vehicle technologies to better fit the needs of novice officers.

1.2. Challenges in current CPMs

There are two major limitations in existing CPMs that need to be addressed. First, current CPMs are limited to evaluating expert performance with rough adjustments to consider novice users, as the purpose of the models was to improve the design of interfaces in human-system interaction. For example, models under the Goals, Operators, Methods, and Selection rules (GOMS) family assume the user is an expert performer. These models include simple ones like Keystroke-Level Models (KLM) and more detailed models, such as Natural GOMS Language (NGOMSL), and Critical Path Method GOMS (CPM-GOMS (John and Gray, 1995)). However, GOMS models suffer from several limitations. For example, the models are only appropriate for routine cognitive tasks (i.e., the user knows exactly what to do in the task situation) and represent expert task performance without errors or have very rough estimations to account for novice users (e.g., use of mental operators "M" in KLM models for novice users) (John and Kieras, 1996).

More advanced CPMs such as Executive Process-Interactive Control (EPIC) (Kieras and Meyer, 1997), Queueing Network – Model Human Processor (QN-MHP) (Liu et al., 2006b), and Adaptive Control of Thought (ACT-R) (Anderson et al., 1997) provided the capability to model parallel activities, included detailed sensory inputs and outputs, and were implemented in different software packages such as LISP, C++, VBA, Cogulator, and CogTool. Some other models such as the State, Operator, And Result (SOAR) (Laird et al., 1991) model does not clearly state whether they have the capability to model novices (Laird and Congdon, 2015). Among those advanced models, the QN family of models and ACT-R have the capability to model novices' performance for some domains. For example, the QN-MHP accounts for the effect of age differences on mental workload by using an age factor (A) representing young vs. older adults. In addition, ACT-R was used to investigate the characteristics of novices' collision avoidance braking behavior (Cao et al., 2014; Zhang et al., 2022). ACT-R was advanced to account for multitasking behavior such as novices' text entry performance using cell phone keypads (Das and Stuerzlinger, 2007). The model predicted the amount of time that unskilled users spent finding a key on a keypad and pressing it repeatedly. The QN approach (Liu et al., 2006a) has been improved with some error modeling capabilities such as a wrongly processed entity or character (Wu and Liu, 2008) and errors in numerical typing (Lin and Wu, 2012). However, these errors were meant to account for experts potentially committing errors due to their cognitive workload rather than to model novices' perceptual, cognitive, and motor characteristics.

ACT-R has also been widely studied for surface transportation domain. Like, QN family of models, ACT-R can model driving performance (Salvucci and Gray, 2004; Salvucci et al., 2007) and interaction with in-vehicle technologies (Salvucci, 2005; Salvucci et al., 2004, 2005; Salvucci and Macuga, 2002). However, the models were not developed for novices. In another ACT-R study, researchers manipulated the model's knowledge and strategies and implemented different production rules for novice and expert models (Cao et al., 2014). However, the objective of this manipulation was to investigate the effect of driving experience on collision avoidance braking behavior. There has not been any study that used advanced CPM techniques to model novices' performance when they are interacting with in-vehicle technologies.

Another major limitation of current CPMs is the accessibility of the models and the difficulty associated with using them for beginners. Some of the developed models (e.g., SOAR and EPIC (Kieras and Meyer, 1997)) are difficult to use for practitioners and experts in domains other than the domain they were originally created for even though there is detailed documentation for the methods (Kieras, 1999). In addition, there are very limited resources such as well-structured manuals, guidelines, or tutorials for learning the basics of each modeling language. An exception to this rule is ACT-R, which has been continuously updated with manuals and tutorials available on its website (http://act-r.psy.cmu.edu/software/) and Github (https://github.com/HOMlab/QN-ACTR-Release). Unfortunately, there is no similar information available for other models except for the material provided in published articles which is not sufficient for beginners to learn the modeling logic. This issue can make the learning process difficult for analysts and practitioners because they might not understand the details of each model (e.g., parameters, inputs, outputs) or whether the method is appropriate to use for their specific application in the first place. Only some GOMS studies share the full source or pseudo codes of the model in the appendix of their manuscripts (Manes, 1997; Paelke, 1993; Schneegaβ et al., 2011). Overall, there is limited information available to allow researchers to replicate the findings of these models, which is essential for their validation (Al Seraj et al., 2018; Wilson et al., 2019). To help alleviate this issue, reproducible practices are needed by emerging technologies such as dynamic document generation tools (e.g., R Markdown), version control and code/data sharing platforms (e.g., Github), and containerization technology (e.g., Docker).

1.3. Differences between novices and experts

A critical difference between novice and expert drivers is the level of cognitive workload (CW) they experience while driving. CW can be defined as "the relation between the function describing mental resources demanded by a task and those resources available to be supplied by the human operator" (Parasuraman et al., 2008, pp. 145–146). Novices tend to look through many chunks of data to find what they need while experts are able to filter extraneous information and find the specific chunks they need more quickly than novices (Carmichael et al., 2010; Sharif et al., 2012). The frequency of saccades and fixations for novices are higher than experts and it takes more time for novices to detect anomalies on roads than experts (Kundel and Nodine, 1975). This is because novices are usually inclined to fixate on visually salient information before focusing on useful semantic information while experts directly focus on semantic information (Sharif et al., 2012). Regarding cognitive processes, experts can deal with more parallel cognitive actions than novices (Kavakli and Gero, 2003). Concerning memory, experts have advantages in chunking ability (Kavakli and Gero, 2003) and the amount of information stored in long-term memory (Sohn and Doane, 2003). For example, expert chess players can easily find and remember positions as familiar configurations or chunks of pieces that they encountered previously (Chase and Simon, 1973). With regards to motor aspects, reaction times for novices are longer compared to experts (Hick, 1952; Hyman, 1953). Experts exhibit fewer motor operators both at the level of central neural programming and subsequent motor unit activation (Davids et al., 2006; McCaskie et al., 2011; Milton et al., 2007). This is evident by novices demonstrating higher engagement in cognitive activity than experts to execute a motor skill that they are learning (McCaskie et al., 2011).

In the surface transportation domain in particular, the CW of novices is significantly correlated with their reduced task performance in high-demand driving conditions (Drummond, 1989). Novices scan more frequently for hazards on the roads (Underwood, 2007) and must put conscious effort into their steering and speed control to avoid road hazards while experts have 'hazard-avoidance schemas' that can be executed in need. This is irrespective of the familiarity of the road and does not imply that expert drivers always avoid hazards, merely that they can respond to adverse situations faster. Novices also lack schema, experiences, and relevant rules of behaviors to effectively complete their tasks in a vehicle (Borowsky et al., 2008). In addition, they have inadequate situation awareness, (McKenna and Crick, 1994) exhibited by shorter glances and longer response times on the phone while driving (Smiley et al., 2007).

1.4. Research objective

The objective of this study was to develop a cognitive performance model for novice law enforcement officers (N-CPM) to model their performance (in terms of the perceptual, cognitive, and motor demands, and task completion time) while interacting with in-vehicle technology. N-CPM is an extension of the original CPM-GOMS model which can be used to model novice behavior. To validate the model, we conducted a ride-along study with novice law enforcement officers (nLEOs) as police operations create high-demand driving conditions for nLEOs (Shahini et al., 2020; Zahabi and Kaber, 2018b). Since this was a naturalistic driving study and officers were on duty, we could not control the tasks that officers performed and instead we identified the most common tasks performed among all participants based on our observation. In addition to the ride-along study, a benchmark model was developed using the CPM-GOMS method and ACT-R working memory formulation in Cogulator software (Estes, 2017) to be compared with the developed N-CPM.

2. Method

The N-CPM is developed to model nLEOs' interaction with in-vehicle technologies. The fundamental human behavior principles in N-CPM are based on Model Human Processor (MHP) (Card and Newell, 1986), CPM-GOMS, and working memory structure in ACT-R. The GOMS family of models have been successful in modeling driver interaction with in-vehicle technology in part because of its simplicity due to being based on MHP (Lee et al., 2019; Liu, 2019; Park et al., 2020; Park and Zahabi, 2022; Purucker et al., 2017; Yang et al., 2019; Zahabi, 2017; Zahabi et al., 2019a). However, GOMS is limited when it comes to providing information regarding working memory process during a task. Therefore, we used the formulations for working memory activation and decay based on ACT-R (which is similar to the logic used in Cogulator software) to address the limitations of GOMS.

We initially considered using QN-MHP or ACT-R as a basis for N-CPM, as our recent literature review study found that these models were frequently used as human performance models in the surface transportation domain (Park and Zahabi, 2022). However, due to the following reasons, we decided to use MHP/CPM-GOMS instead of QN-MHP. First, the models that used novice drivers were focused on estimating the primary task performance (e.g., braking) (Cao et al., 2014; Zhang et al., 2022) and not drivers' interaction with in-vehicle technologies, which was the focus of our study. Similarly, ACT-R was advanced in specific applications to account for novices' text entry performance using cell phone keypads (Das and Stuerzlinger, 2007) and was not used to model drivers' interaction with in-vehicle technology. Second, implementing the logic in ACT-R requires development of numerous production rules and modifications to its internal logic (e.g., Fitts' law). Our goal was to provide a simple model that can be used by analysts without expertise in human performance modeling. Due to the modeling complexity of both ACT-R and QN-MHP, we decided to use MHP/COM-GOMS models as a basis for N-CPM.

The language used to develop the scenarios followed the programming language used in the Cogulator software. The outcomes from N-CPM are task completion time (TCT) based on the summation of the time of all perceptual, cognitive, and motor operators for a given task while considering parallel activities. Additionally, the model provides the number of memory chunks (MC), and perceptual/motor/cognitive operators used during the task. N-CPM also has the capability to develop a scenario or cognitive model (defined as a series of perceptual/cognitive/motor operators, methods, and selection rules used by an individual to accomplish a specific goal) using a graphical user interface (GUI). The developed N-CPM can be downloaded from Github (https://github.com/hsilab/ncpm_v1.0) and installed in Rstudio as a package. The R version used for the model development was 4.0.5. There are some prerequisite packages that must be installed before using this package, including devtools (Wickham and Chang, 2016), ellipsis (Wickham, 2021), vctrs (Wickham et al., 2022), shiny (Rstudio, 2014), Rcpp (Eddelbuettel et al., 2022), and skimr (Waring et al., 2022).

2.1. Operators

To develop the N-CPM, a logic was needed to differentiate the number of operators between novices and experts. Operators in N-CPM refer to basic actions of users, which can be categorized as perceptual (e.g., vision), cognitive (e.g., memory retrieval), or motor (e.g., moving hands) operators. Previous models were developed based on the assumption that experts do not commit errors. The experts' performance in N-CPM has the same assumption. Therefore, the time to execute each operator for experts came from previous studies and was described in the glossary tab of the model GUI (Card and Newell, 1986; John and Gray, 1995; Kieras, 1997; Kim and Myung, 2016; Nyström, 2018; Park and Zahabi, 2021). To adjust the operators to account for novices' behavior, we used the Rasmussen (1983)'s Skill-Rule-Knowledge (SRK) based human behavior framework and prior studies that compared the performance of novices and experts.

Rasmussen (1983) classified types of human behaviors in three levels. Skill-based behavior (SBB) occurs in a known context and the skill-based tasks are performed without conscious attention or control (Rasmussen, 1983; Reason, 1990). Skill-based responses are generally initiated by some specific event, such as an experienced driver stopping at a stop sign. Several police in-vehicle tasks require only skill-based interactions such as "picking up the phone." At the skill-based level, only perceptual operators were inflated to account for differences in perception of novices and experts as novices can have additional pursuit eye movements than experienced drivers (Mourant and Rockwell, 1972). Skill-based tasks require a minimal number of cognitive operators (Rasmussen, 1983; Reason, 1990). However, even for minimal attention, at least two cognitive operators are required: initiation and verification (Embrey, 2005).

Rule-based behavior (RBB) occurs in familiar contexts with environmental information being sensed as signs (i.e., to modify predetermined actions based on convention or previous knowledge). Rule-based tasks are performed based on stored rules or procedures. Rules can be thought of as "if-then" associations between inputs and appropriate actions (Embrey, 2005). Knowledge-based behavior (KBB) occurs in unfamiliar contexts, with the environmental information being sensed as symbols (i.e., information used to predict or explain non-familiar situations). Conscious problem solving and planning is required in KBB, which requires significantly more attentional resources compared to the skill- and rule-based behaviors. Reason (1990) stated that more cognitive activities are needed for knowledge-based tasks as compared to skill-based tasks. Examples of knowledge-based tasks in the vehicle are those secondary tasks that require multiple interactions with the technology and include perceptual and motor operators as well as additional cognitive operators.

We conducted a literature review to quantify the differences between novices and experts and used that to modify the operators to account for novice performance in N-CPM. However, using this approach might have some limitations as not all previous studies were focused on novices' interaction with in-vehicle technologies. We assumed that novices can have 1 or 2 more perceptual operators than experts when performing the same task based on Law et al. (2004). This study compared the frequency of fixations between experts (8.9%) and novices (26.7%) during a surgery task and found that the number of fixations for novices was three times higher than that of experts.

Another assumption is that novices use 2 or 3 times more cognitive operators when performing a task than experts. This assumption was based on a study by Kavakli and Gero (2003) that found novices had 2.8 times as many segments (i.e., cognitive actions that appear to occur simultaneously) as expert designers. Related to this, McCaskie et al. (2011) found that novices engage in more cognitive activities than experts while learning motor skills (McCaskie et al., 2011).

Some studies found that novices use 1 or 2 more motor operators than experts due to the errors made while performing a task (Goode et al., 1998; Milton et al., 2007; Wiedenbeck, 1985). For example, in sports, the error rate in performing motor activities for experts was 7.7%, while the error rate for novices was 11.9% (Goode et al., 1998; Milton et al., 2007). Experts are faster and more accurate in recognizing patterns and are better able to plan actions in advance (Goode et al., 1998; Milton et al., 2007). Experts also perceive kinematic information better, produce more consistent and adaptable movement patterns, use domain-specific information more effectively and efficiently, and have superior procedural and declarative knowledge (Goode et al., 1998; Milton et al., 2007).

A summary of the operators in N-CPM is shown in Table 1. The inflated operators for novices are multiple instances of the same activity (e.g., perceptual operators can range from 1–4). The N-CPM randomly select a number within this range in each run. For example, novices might need to look at a display more than experts to confirm the information. However, due to individual differences, some novices might have similar behavior as experts. Table 1 can be used as a general rule-of-thumb regarding the differences in the number of operators between experts and novices and is not intended to be used for a specific domain. The adjustment to the noise parameter is discussed in the following section.

2.2. Working memory

Activation is a degree of association between previous experiences and current context which describes whether a chunk will be helpful at any given moment (Bothell, 2017). Chunks are the elements of declarative knowledge in the ACT-R theory and are used to communicate information among modules through the buffer (Bothell, 2020). The activation of a memory trace is calculated using the following equation (Altmann and Schunn, 2019; Estes, 2015):

$$Activation = \ln\left(\frac{n}{\sqrt{T}}\right) \tag{1}$$

where n is the number of times that chunk is rehearsed, and T is the total time the trace is held in memory (or age of the item). The number of rehearsals refers to the familiarity of a chunk. The default value of this parameter was set as 3, a plausible level of rehearsal that exhibited the best overall fit (Estes, 2015). However, for the information from long-term memory (LTM) or recall information, the number of rehearsals was set to 10 to indicate that a chunk from LTM is difficult to be forgotten. (Estes, 2021). This number can be updated while the model is running. For example, if the number of rehearsals increases, the activation also increases, leading to higher recall probability.

In order to mimic the division of activation across all working memory chunks, the activation was reduced as a function of the number of chunks in the problem span based on the logic in Cogulator software (Estes, 2021). Therefore, the divided activation in N-CPM was calculated as shown in Equation (2) (Estes, 2015).

$$Divided\ activation = Activation + \frac{1}{stack\ depth} - 1 \tag{2}$$

"Stack depth" is the number of chunks the activation must be divided among. For example, if a chunk was added as a third chunk in the memory, the stack depth becomes 3. The idea of limited activation source pools and their distribution among all the chunks held in working memory has been previously documented in the literature (Anderson et al., 1996). Equations (1) and (2) allowed the N-CPM to model a relationship between the number of chunks to be memorized (e.g., visual pattern span), decay over time, and a subjective rating of mental demand.

Memory load was defined as the overall occupancy of chunks in the entire task and was calculated using Equation (3), which divides the summation of the duration of all chunks by the total task duration (Estes, 2015, 2021).

$$Memory\ load = \frac{\sum_{i=1}^{7} (chunk\ duration)_i}{total\ task\ duration} \tag{3}$$

2.3. Recall probability

Based on the activation calculated from Equation (2), recallability was calculated as shown in Equation (4) (Dehban et al., 2015; Estes, 2021).

$$P_i = \frac{1}{1 + e^{\frac{\tau - A_i}{s}}} \tag{4}$$

In this equation, τ is a threshold to forget the chunk (−1), s is the noise or the variance from one scenario to another (Bothell, 2017), which is set to 0.8 based on Estes and Masalonis (2003), and A_i is the activation value from equation (1). Finally, P_i refers to the recall probability of the ith chunk. The default value of threshold and noise came from the Cogulator software as the model assumes the user is an expert. The N-CPM assumes that the speed at which memory decays for novices will be faster than experts by adjusting the noise parameter in equation (4). As observed in expert's behavior, continued training leads to a reduction in reaction time and errors, and eventual power law shape of

Table 1
Summary of parameters for novices and experts in N-CPM.

Parameters		Level of expertise			References
		Experts	Novices		
			Skill-based behaviors	Rule-based and Knowledge-based behaviors	
Perceptual operators	Inflation factor	1	1∼2	1∼4	Law et al. (2004)
Cognitive operators	Inflation factor	1	1	1∼3	Kavakli and Gero (2003)
	Noise in Recall probability	0.2	0.8		Dehban et al. (2015); Di Nota and Huhta (2019); Estes (2021); Estes (2015);
Motor operators	Inflation factor	1	1	1∼4	McCaskie et al. (2011)

performance to execute now-automatized behaviors (Fitts and Posner, 1967; Ritter and Schooler, 2001). In addition, expert sensorimotor networks facilitate decision-making, performance, and novel motor learning that is faster and more accurate than novices. Thus, N-CPM assumes the noise (or variance between the scenarios) parameter for experts is smaller than that of novices (s for experts = 0.2; s for novices = 0.8).

2.4. How to use N-CPM?

When analysts run the N-CPM after installing the package, they first see the screen displayed in Fig. 1. Analysts can directly develop a scenario or load a developed scenario if one is already prepared in CSV format.

To develop a scenario, analysts can select operators from the screen after clicking on the “Develop a Scenario” tab. Chunk names and descriptions of each line can also be added. The created scenario/code will then be shown on the right side of the window. Fig. 2 illustrates an example of the “Develop a Scenario” tab. Note the example code on the right side of the screen.

The scenario can be generated from “Scenario development” tab (Fig. 1), notepad, or Cogulator. For example, in Fig. 2, the goal of the scenario is to check a vehicle plate number. A memory chunk used during the task is modeled with a bracket (”<” and “>“). For example, in Fig. 2, “<HZO>” was used as one chunk because it was the first half of the plate number and “<01 K>” is another chunk which is referring to the second part of the plate number.

In the “Results Summary” tab (Fig. 3), all the outcomes for novices are presented including the task completion time, memory load, and number of operators. In addition, the information contained in each chunk is presented in a table format. In this table, there are columns for the number of rehearsals (the number of accesses on that specific chunk during the task), activation (calculated from Equations (1) and (2)), and probability of recall (calculated based on Equation (4)).

As soon as the analyst develops a cognitive performance model, N-CPM counts the number of cognitive operators. Then, based on the rule described in Table 1, N-CPM applies the additional operators. All the comparisons between novices and experts are located in the “Novices and Expert Comparison” tab (Fig. 4).

Lastly, the N-CPM provides tutorials for scenario development and editing within the “Help” tab as shown in Fig. 5. The description and duration for all operators including references is found in the “Glossary” tab (Fig. 6).

2.5. Model validation

A ride-along study was conducted with 10 nLEOs (Age: M = 28.6 yrs, SD = 5.64 yrs) (Wozniak et al., 2022). Each participant was observed for a period of at least 3 h. The study protocol was approved by the Texas A&M institutional review board (IRB 2021-0757D). In the following sections, we describe the participant demographic information, study equipment and procedure, and analysis approaches.

All officers were novices, defined as having fewer than five years of experience as a primary patrol officer (Patrol officer experience: M = 2.15 yrs, SD = 1.23 yrs) (Filtness et al., 2013). To control for driving experience being a factor in novice performance, participants were required to have at least 1.5 years of driving experience. The study was conducted during officers’ regular work shifts in mornings and afternoons. Weather during these ride-alongs ranged from sunny to overcast to rainy.

A dash camera was attached to the roof of the vehicle to observe the officer’s interactions with their in-vehicle technology. The study began with participants filling out an informed consent form and a demographic questionnaire. Several in-vehicle devices were used for the patrol mission as shown in Fig. 7. After the data collection, the participant returned to their office. The participant was then provided with a copy of the informed consent form for their reference and thanked for their time.

The Cogulator software which uses the CPM-GOMS method (to model expert performance) as its logic was used as a benchmark model in this study. After researcher 1 finalized task identification and analysis, they created initial benchmark models. Then, researcher 2 reviewed and revised the models. Once researcher 1 and 2 agreed on all the models, researcher 3 reviewed the models. Lastly, researcher 1 and 2 revised the models based on researcher 3’s feedback to reach 100% intercoder agreement.

2.6. Data pre-processing and analysis

Task analysis was conducted based on the recorded videos using VideoPad v6.26 software. Perceptual and motor operators were manually identified from the video analysis and timed to the 33.3 ms (30 frames per second). The video quality was sufficient to see the field of view (windshield) of the driver and the in-vehicle technology. This also allowed for the identification of patterns of repetitive perceptual and motor operators that were incorporated into the N-CPM as shown in Table 1. The timing of the visual operators initiated when the officers’ eyes started moving from the windshield to a target (e.g., MCT). The timing of auditory operators began from the onset of the auditory stimuli. In addition, motor operators were tracked from onset of

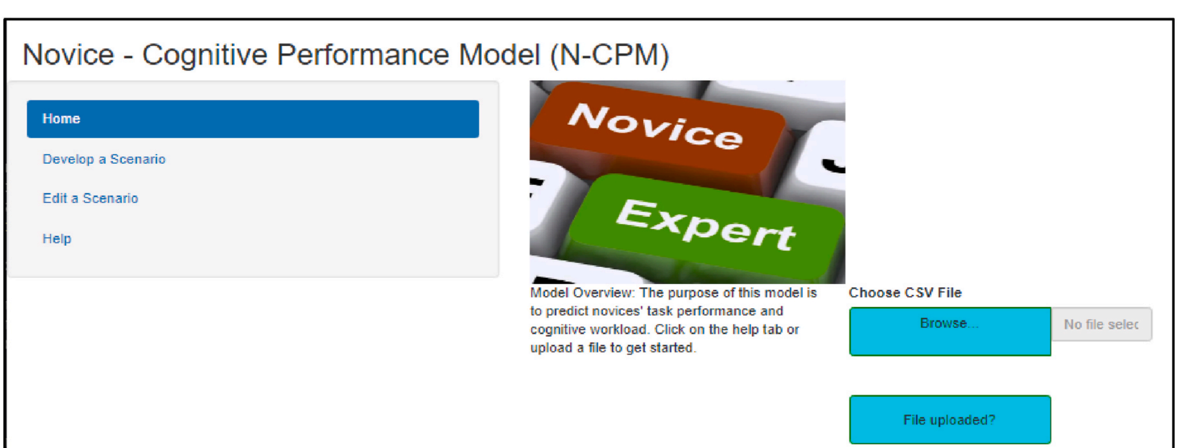


Fig. 1. Initial (Main) screen of N-CPM.

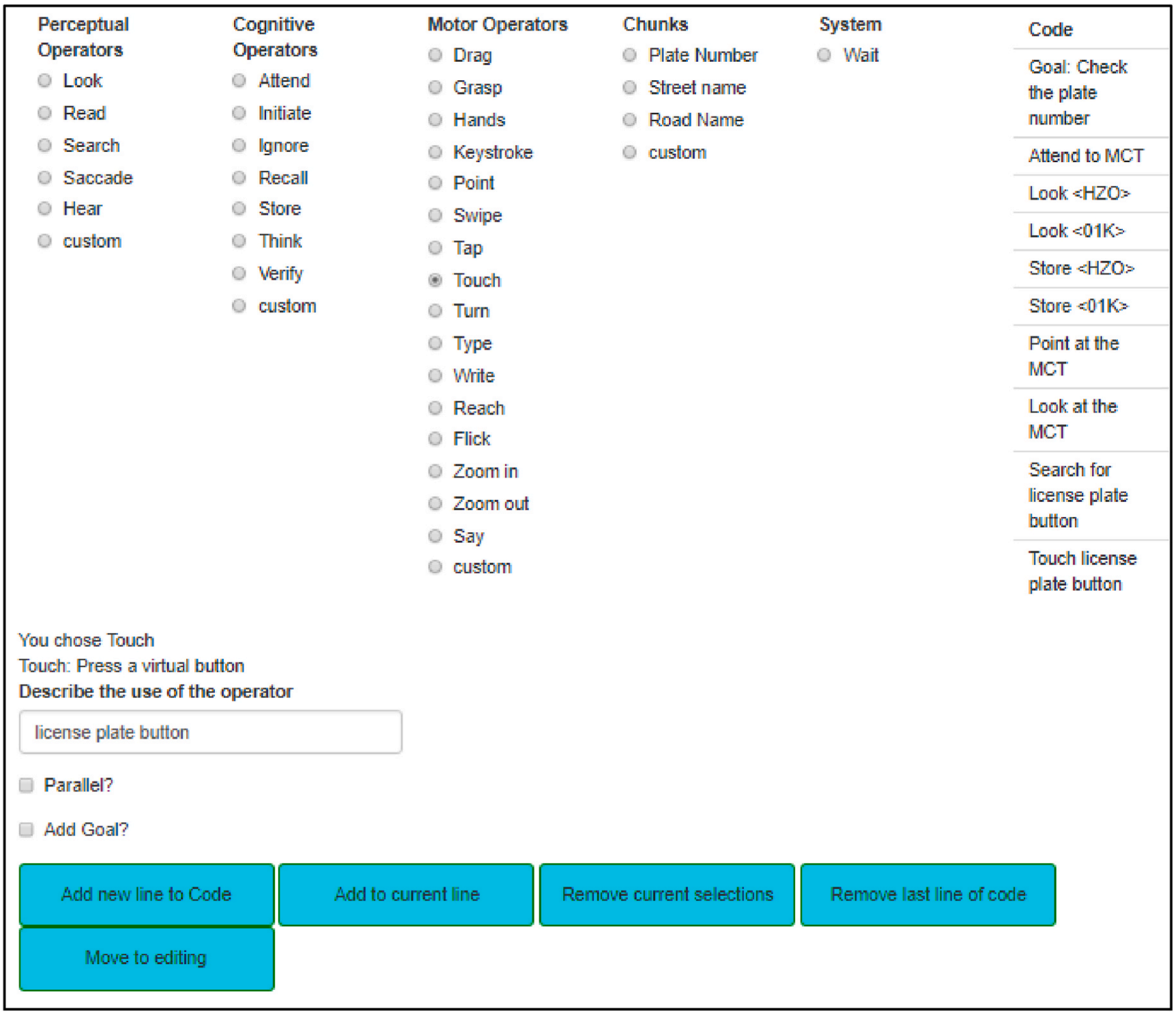


Fig. 2. "Scenario development" screen.

movement to completion.

Based on the task analysis and the most frequent tasks performed during the ride-alongs by the participants, we extracted 16 scenarios as shown in Table 2. These 16 scenarios also represent the most frequently performed activities in police vehicles based on prior studies (Zahabi and Kaber, 2018b; Zahabi et al., 2019a). TCT, the number of perceptual operators (nP), and motor operators (nM) for each scenario were extracted from this process. In addition, the cognitive models for experts were developed after the task analysis as described in the model validation section. From Cogulator, TCT, nP, number of cognitive operators (nC), nM, and memory chunks were calculated for experts. The N-CPM also generated the same metrics for both novices and experts.

Statistical analysis was conducted using JMP 16 and R 4.0.5 software. The statistical models were tested for normality and equality of variance, however none of the assumptions were met for the human data collected from the ride-along observations and the benchmark model data. Therefore, the nonparametric Wilcoxon test was conducted. The *p*-value, *Z* score, and effect size were reported for each test. We used an alpha level of .05 for all statistical tests. There were 70.36 data points on average from each participant ($SD = 43.82$). This means that each participant exhibited about 70 in-vehicle interactions on average. From the tasks' point of view, each task was performed by officers 48.38 times on average ($SD = 61.65$). As this study was a naturalistic ride-along study, we did not limit the number or types of tasks. Furthermore, the normality and equality of the variance assumptions were not met, as each participant was in a different situation. That is, some officers performed the tasks more frequently or exhibited longer task completion times than other officers. Also, the violation of assumptions in the benchmark model data might have occurred due to the deterministic nature of the model and since the outcomes of the models are based on expert performance.

2.7. Hypotheses

Scenarios S2, S5, and S6 were removed from statistical analysis due to the lack of sufficient data points obtained from the video analysis as officers did not typically perform these scenarios across all the ride-along samples. These scenarios are still mentioned because they represent important tasks that officers must complete during their patrol task, even if they are not performed as frequently as some other tasks. The hypotheses shown in Table 3 below were only generated for the TCT measure because all three sources (i.e., N-CPM, benchmark model, and observation data) could generate TCT, while the number of cognitive operators and memory chunks could only be calculated from N-CPM and Cogulator. In total, three hypotheses were formulated as shown in Table 3.

The scenarios were categorized as simple or complex scenarios based on the "15-Second Rule" (Green, 1999, 2008), which is the recommended maximum time for drivers to complete in-vehicle-information-system-related tasks involving visual displays and manual controls. The hypotheses are only focused on complex tasks because simple tasks did not have enough operators and by extension a

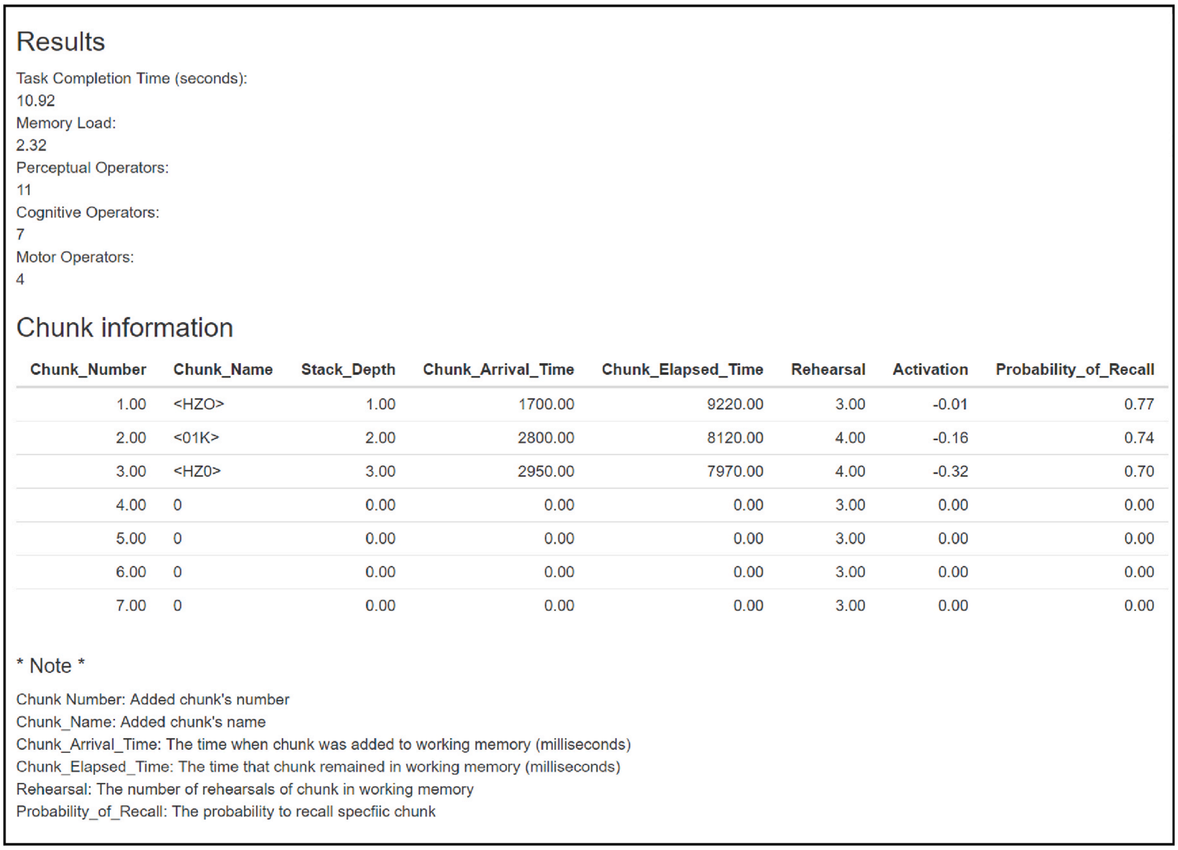


Fig. 3. "Results Summary" screen.

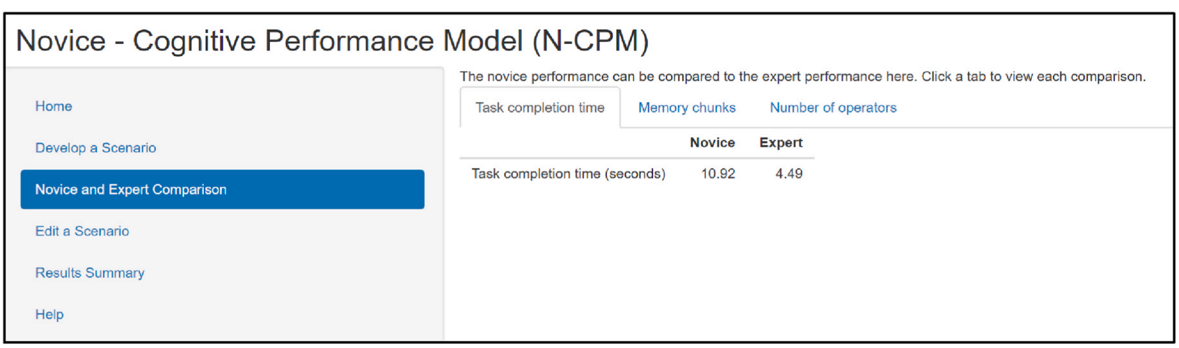


Fig. 4. "Novice and Expert Comparison - Task Completion Time" screen.

high enough TCT for average novice performance to be significantly different from expert performance, meaning that experience was not a significant factor in performing simple tasks. Simple tasks can be identified as tasks that naturally fall under the control of skill-based behavior under Rasmussen's SRK model (Rasmussen, 1983). These tasks can be completed without much conscious input from the participant due to being completed frequently in other contexts that do not have to do with the driving task. Additionally, by definition, more complex tasks engage rule-based and knowledge-based behaviors that create differences between novice and expert drivers, with rule-based complex tasks primarily consisting of tasks that participants are not as familiar with and knowledge-based complex tasks presenting new scenarios that novice drivers are not prepared to address. Therefore, the following hypotheses in Table 3 were formulated around the complex tasks.

3. Results

A summary of the mean results for each task's TCT is shown in Table 4. Table 5 summarizes the results for the number of operators for each task as well as the memory chunks. The number of cognitive operators or memory chunks could not be captured from the observation data because they are implicit variables. Simple task results are included here to demonstrate the differences between these tasks and complex tasks as well as to indicate why the task load on officers is so high while they complete the patrol task. This is also done to indicate that there are several tasks that fall in line with expert performance while highlighting the importance of designing technology around tasks that novices would struggle more than experts. Benchmark models were only run once to obtain a TCT because they only generate a single output due to being based on expert performance. TCT for observation data was based on the average of all performances of the tasks by participants. N-CPM TCT was determined using the average of multiple runs for each scenario. The N-

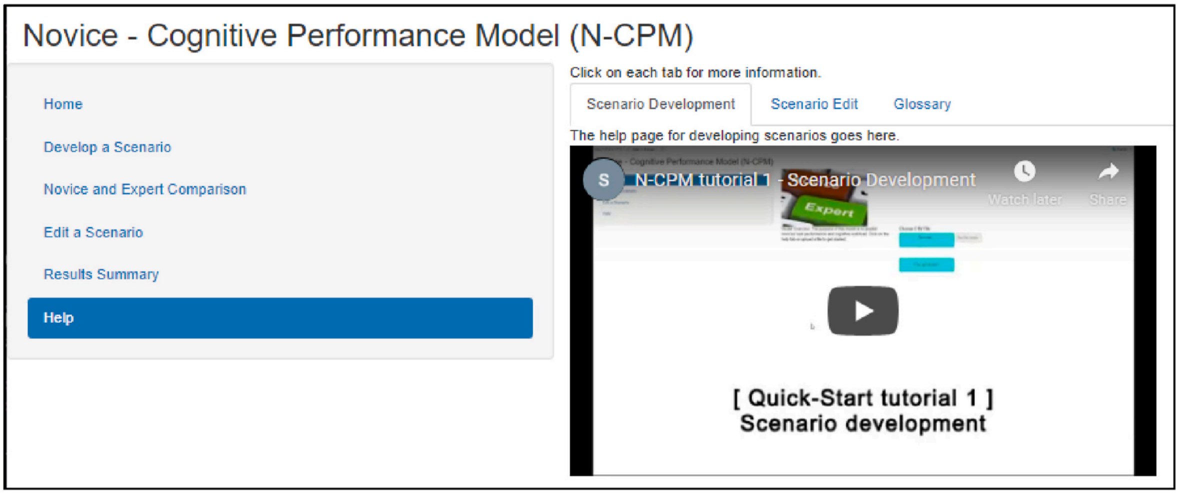
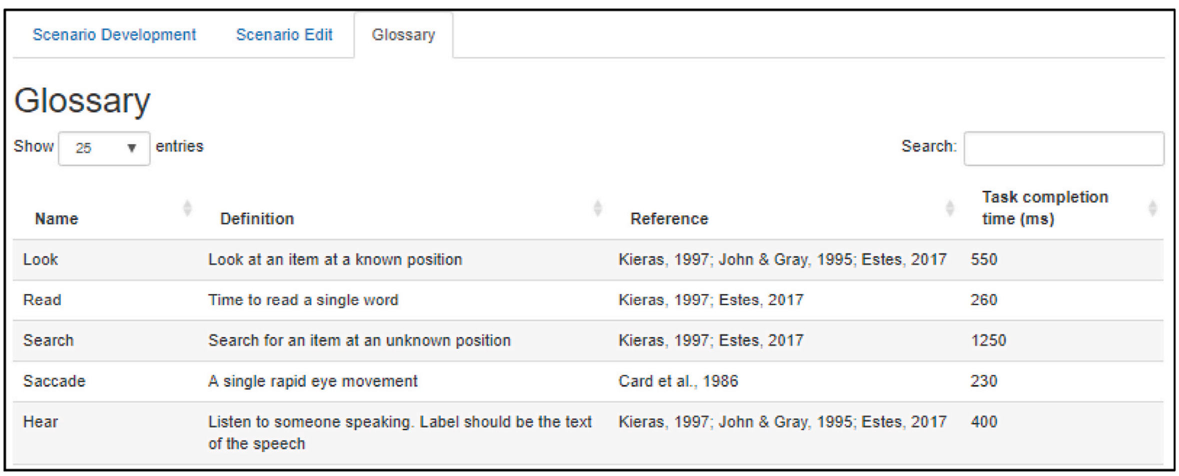


Fig. 5. Help video for "Scenario Development".



Name	Definition	Reference	Task completion time (ms)
Look	Look at an item at a known position	Kieras, 1997; John & Gray, 1995; Estes, 2017	550
Read	Time to read a single word	Kieras, 1997; Estes, 2017	260
Search	Search for an item at an unknown position	Kieras, 1997; Estes, 2017	1250
Saccade	A single rapid eye movement	Card et al., 1986	230
Hear	Listen to someone speaking. Label should be the text of the speech	Kieras, 1997; John & Gray, 1995; Estes, 2017	400

Fig. 6. Glossary.

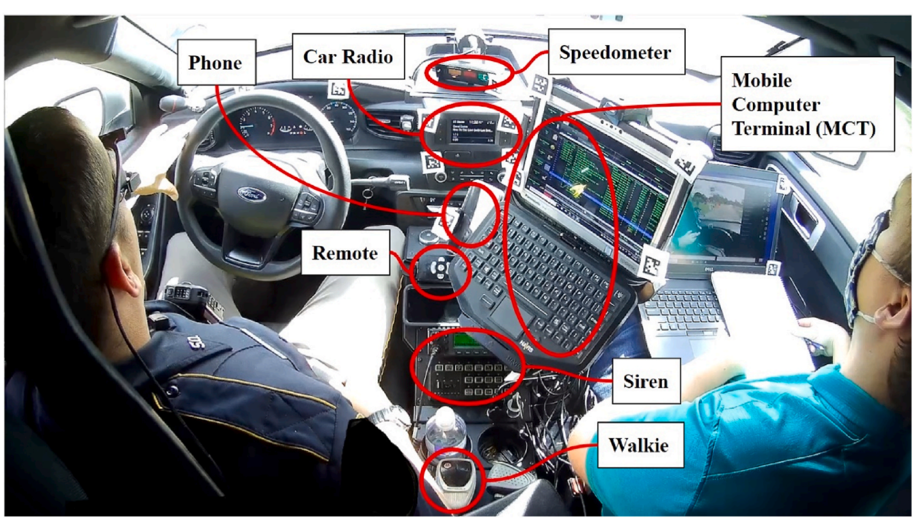


Fig. 7. A dashcam screenshot (ride-along).

CPM model was run multiple times for each scenario depending on the number of observation data for that scenario to have a comparison between the human data and model data.

The results of the hypothesis tests are shown in Table 6. We did not find adequate evidence to reject our third hypothesis (i.e., there is no significant difference between N-CPM outcomes and observation data),

Table 2
Scenario information.

Scenario ID	Title	Description	Device
S1	Make a call	Call someone or answer a call on the phone and have a conversation	Phone
S2	Move a phone	Move a phone from one location to another within the vehicle	Phone
S3	Read a notification	Look at a phone notification without opening the phone	Phone
S4	Read information	Open the phone and read some information on it	Phone
S5	Reject a call	Reject an incoming phone call	Phone
S6	Respond to text	Read and respond to a text message on a phone	Phone
S7	Use map app	Use the map application on a phone to input a destination	Phone
S8	Use touchscreen	Use a smartwatch or other screen in the vehicle to read or find information	Screen
S9	Change music	Change the music on a car radio	Radio
S10	Press button	Press a button on a remote to check the speed of nearby cars.	Remote
S11	Use siren	Press a button to turn on/off police sirens	Siren
S12	Communicate	Have a conversation on a walkie	Walkie
S13	Pick up	Pick up a walkie	Walkie
S14	Talk on	Speak into a walkie and hear a response once	Walkie
S15	Read information	Passive-level MCT task that involves reading information on the MCT without directly interacting with it	MCT
S16	Search and read information	Active-level MCT task that involves typing or tapping on the MCT to find or input necessary information	MCT

*S2, S5, and S6 were removed from statistical analysis due to the lack of data samples.

Table 3
Hypotheses.

Hypotheses (H)
H1: TCT of N-CPM would be significantly higher than the benchmark model for complex tasks
H2: TCT of observation would be significantly higher than the benchmark model for complex tasks
H3: TCT of observation would be similar to the N-CPM for complex tasks

Table 4
Descriptive statistics of the TCT results for each scenario.

Scenario Complexity	Source			
	ID	TCT (s)		
		Observation	N-CPM	Benchmark model
Complex	S1	80.9	90.28	59.3
Complex	S7	53.7	51.65	15.4
Complex	S12	25.2	31.7	15.9
Complex	S16	30.0	33.42	11.5
Simple	S3	9.87	7.63	5.2
Simple	S4	6.91	7.79	5.04
Simple	S8	7.01	7.29	5.71
Simple	S9	3.91	4.03	3.1
Simple	S10	2.2	2.12	1.8
Simple	S11	6.4	6.07	5.1
Simple	S13	4.08	4.02	4.02
Simple	S14	11.0	9.49	8
Simple	S15	8.39	6.34	5

Note: TCT = Task completion time.

except in S12 (i.e., have a conversation on a walkie). This means that the N-CPM was able to model the performance of novices with reasonable accuracy. Regarding H1, there were significant differences between N-CPM and benchmark model gathered from Cogulator for both simple and complex tasks. This is likely because the standard deviation of the benchmark results generated from Cogulator is smaller than the variation in N-CPM (due to the inflation factors listed in Table 1) and observation data. These findings imply that the N-CPM results are significantly different from expert performance, and when combined with the findings from H3 implies that N-CPM more accurately models novice performance than expert performance for these tasks. The results for the simple tasks are included to be comprehensive but are not relevant to the findings of the hypothesis tests.

To further verify the model, the root mean square error (RMSE) and R-squared were calculated (Wu, 2018). RMSE for TCT between the observation data and N-CPM was 3.49s with an R-squared value of 0.99, while RMSE of TCT between observation and the benchmark model was 13.62s, with R-squared value of 0.88. Between the N-CPM and the benchmark model, RMSE was 15.26s and R-squared was 0.88.

4. Discussion

The results demonstrated several merits of N-CPM. We could not find adequate evidence to reject this hypothesis that the observation data were similar to N-CPM outcomes. Furthermore, both the N-CPM and observation results were significantly larger than the benchmark model outcomes for most complex tasks. For most simple tasks, the N-CPM and benchmark model outcomes were not significantly different from the observation data. Hypothesis 2 (H2) was only rejected for one complex scenario (i.e., S1) while H1 was supported for all complex scenarios. The N-CPM TCT results for simple tasks were also significantly higher than the benchmark data. Although the tasks were simple, benchmark models based on expert performance produced smaller standard deviations compared to the N-CPM results. Despite the simplicity of the task allowing novice performance modeled by the N-CPM to mimic expert performance, the lack of consistency compared to the expert benchmark performance was likely enough for the simple tasks to have a significantly higher N-CPM TCT compared to the benchmark TCT estimates. The merits of the N-CPM over a benchmark for novice performance are primarily demonstrated by the results of H3 and H1 for complex tasks. Despite the higher standard deviation in these complex tasks due to the variability in how they were approached by novices, the findings of N-CPM for complex tasks were found to not be significantly different from the observation data while also being significantly different from the benchmark data. This means that the N-CPM TCT estimates were more accurate at representing observation TCTs as opposed to benchmark TCTs.

We intentionally placed several scenarios under each hypothesis to try to generalize the results of hypotheses testing across several in-vehicle technologies for nLEOs. However, H2 was not supported in one scenario (e.g., making a phone call). One reason for this was likely the small sample size (i.e., only 7 observations) and a consequentially higher than expected standard deviation of TCT for these tasks from observation. Tasks that included communication had much higher variation in duration which led to some significant differences not related to the complexity of the task. For example, H3 was rejected in one complex scenario (S12). One possible explanation for the difference between the observation data and the N-CPM outcomes in S12 (i.e., Have a conversation on a walkie) was the variations existed in the human data due to differences in the length of conversation for each participant. Overall, N-CPM could model novices' cognitive performance accurately for most scenarios and performed better than the benchmark model in complex scenarios. This was predictable because the benchmark model is developed for experts.

Other metrics besides TCT were also in line with the findings of previous studies (Chase and Simon, 1973; Sohn and Doane, 2003). The N-CPM consistently outputs more perceptual and motor operators than the benchmark model for most tasks, which is closer to the number of operators seen in observation tasks (Table 5). It was not possible to

Table 5
Descriptive statistics of the results for each scenario.

ID (Complexity)	Source									
	Observation		N-CPM				Benchmark model			
	nP	nM	nC	nP	nM	MC	nC	nP	nM	MC
S1 (C)	7.29	18.43	28.80	51.41	20.55	0.69	6	7	12	0.8
S7 (C)	16.1	12.79	24.78	17.23	21.59	0.96	6	7	9	0.96
S12 (C)	0.89	7.64	20	11	13.58	1.35	4	3	7	1.63
S16 (C)	10.9	11.76	16.57	27.84	7.77	0.92	4	11	4	0.95
S3 (S)	2.64	5.27	2.39	6.18	6.38	0.31	2	3	4	0.31
S4 (S)	3.2	4.53	2.62	1.54	8.82	0.90	2	1	6	0.89
S8 (S)	3.25	4.63	2.83	1.18	8.83	0.10	2	1	5	0.09
S9 (S)	1.5	2.42	2.44	0	2.69	0.79	2	0	2	0.82
S10 (S)	0.71	2.87	0	1.53	2	0	0	1	3	0
S11 (S)	0.89	3.22	2.24	1.61	3.98	0	2	1	3	0
S13 (S)	0.12	2.73	0.98	0	2.99	0	1	0	3	0
S14 (S)	0.73	5.32	2,96	1.96	7	0.95	2	1	5	0.94
S15 (S)	3.68	4.02	2.55	14.08	0	2.6	2	5	9	0.58

Note: (C) = Complex task, (S) = Simple task, nC = Number of cognitive operators, nP = Number of perceptual operators, nM = Number of motor operators, MC = memory chunks.

exactly match the number of operators from N-CPM with the operators captured from observations as the standard deviation of the number of operators from observations is large. This large standard deviation stems from the wide variety within each task that nLEOs have to contend with in each ride-along, increasing the opportunity for errors due to lower experience. This opportunity for errors was more present in the nLEOs recruited given that they had less than 5 years of experience and were more prone to making erroneous decisions. Increasing mistakes also increased the number of operators used for each task, which was reflected in the design of the N-CPM. For complex tasks, the magnitude of inflated operators in N-CPM was larger than that of simple tasks, as complex tasks require more human information processing which is in line with previous studies (Fincham, 2005; Oyewole et al., 2011).

There are several benefits of using N-CPM for researchers. First, there is no need to use a separate code to model novice behavior. The N-CPM has an algorithm to estimate the performance of both novices and experts. This could make it easier for analysts to develop scenarios because they do not have to model repetitive interactions of novices, as the model already accounts for this. Furthermore, analysts could compare the outcomes of the model for novices and experts. Second, N-CPM codes/scenarios can also be run in Cogulator. Analysts can develop a scenario using the N-CPM GUI and paste it in Cogulator. Also, the N-CPM can run the model with a Cogulator scenario file (Microsoft CSV file). This can greatly improve the scenario development process, as sometimes experienced analysts prefer working with spreadsheets rather than the GUI. N-CPM provides a venue (the “Scenario development” tab) for analysts who are not familiar with cognitive performance modeling to create models (Fig. 2). Third, N-CPM is the first CPM that was developed and released in R package format to Github (https://github.com/hsilab/ncpm_v1.0). The outcome of the N-CPM can also easily be integrated with machine learning algorithms for future studies (Zahabi et al., 2020b). Although N-CPM runs on R, its GUI was developed with the Shiny package, which makes the N-CPM accessible to anyone with basic knowledge of R.

4.1. Practical implications

The N-CPM can be applied to improve nLEO’s patrol mission performance through improving in-vehicle technology interaction and adaptive training based on their projected performance to reduce their workload and driving distraction. Potential examples of this include teaching officers on how to reduce the number of interactions with in-vehicle technologies that are linked to high workload by the N-CPM findings.

The outcomes from the N-CPM can also be used to classify nLEO’s cognitive status in real-time with machine learning algorithms. This is most useful for identifying situations where the workload of nLEOs is the highest while operating their vehicle and recommending changes on how the in-vehicle technology should be designed or interacted with. It can also support the design of MCTs with regards to information presentation or prioritization. For example, having a more adaptive design for the MCT screen would reduce CW if another task requiring high CW such as making a phone call or sending a text message had to be completed simultaneously. In addition, N-CPM can be used to provide adaptive training for nLEOs (Zahabi and Abdul Razak, 2020; Zahabi et al., 2020b), in which the problem, stimulus, or tasks dynamically change based on trainees’ performance (Kelley, 1969; Kelley, 1969). Previous adaptive training studies did not support modeling the cognitive process of novices or their prior knowledge that is essential for learning (Kozhevnikov, 2007). This can now be accomplished with the N-CPM. Observations from prior work conducted on adaptive driving simulation-based training and adaptive virtual reality-based training indicate that the implementation of these training protocols can enhance officer performance and reduce workload in high-demand situations. An example of this would be having an MCT that provides only the most necessary and salient information when it is detected that the officer is conducting high-cognitive-workload secondary tasks while driving. Simulating these scenarios will allow for the development of technology that can reduce officer cognitive workload. The N-CPM can be used to support this type of training by generating models of the perceptual, cognitive, and motor demands for different driving scenarios to be used as offline input measures for an adaptive driving simulation-based training system. N-CPM can also be applied to other human-machine interaction applications. For example, it can be used to test the usability of a newly designed mobile application for novice users. N-CPM is a quick solution (due to the use of a GUI) to compare performance of novices and experts during the early stages of the design and development process and once the analyst has access to initial prototypes.

4.2. Limitations

The major limitation of the N-CPM is that it did not delve into the fundamental human information processing principles with advanced quantification approaches. In other words, the model did not incorporate humans’ brain activities as we analyzed the data based on the video recordings. The N-CPM uses the parameters from MHP and CPM-GOMS. The current model does not reflect the characteristics of novices’ memory decaying speed. Therefore, in the next version of the N-CPM, the model needs to be improved to include fundamental psychological principals for novices as other CPMs do (i.e., ACT-R or QN family of

Table 6
Summary of hypotheses tests.

Hypothesis	Statistical analysis results (p-value, Z-score Z, effect size r)		Hypothesis test Result
	Simple Tasks	Complex Tasks	
H1. TCT of N-CPM would be significantly higher than benchmark for complex tasks	S3: $p = .002$, $Z = 3.50$, $r = 0.88$ S4. $p = .0012$, $Z = 3.83$, $r = 0.86$ S8. $p = .004$, $Z = 2.67$, $r = 0.88$ S9. $p = .005$, $Z = 5.09$, $r = 0.75$ S10. $p < .001$, $Z = 4.27$, $r = 0.53$ S11. $p = .003$, $Z = 3.29$, $r = 0.88$ S13. $p < .001$, $Z = 4.82$, $r = 0.71$ S14. $p < .001$, $Z = 1.75$, $r = 0.32$ S15. $p < .001$, $Z = 5.77$, $r = 0.43$	S1. $p = .006$, $Z = 3.02$, $r = 0.87$ S7. $p < .001$, $Z = 3.76$, $r = 0.86$ S12. $p < .001$, $Z = 6.22$, $r = 0.92$ S16. $p < .001$, $Z = 5.27$, $r = 0.43$	**Supported**
H2. TCT of observation would be significantly higher than the benchmark for complex tasks	S4. $p = .12$, $Z = 1.53$, $r = 0.28$ S8. $p = .43$, $Z = 0.58$, $r = 0.14$ S9. $p = .34$, $Z = 0.95$, $r = 0.13$ S10. $p = .33$, $Z = -1.80$, $r = 0.19$ S11. $p = .72$, $Z = -0.43$, $r = 0.10$ S13. $p = .57$, $Z = -0.57$, $r = 0.06$	S1. $p = .70$, $Z = 0.37$, $r = 0.10$	**Rejected**
	S3. $p = .04$, $Z = 1.77$, $r = 0.38$ S14. $p < .001$, $Z = 3.71$, $r = 0.57$ S15. $p < .001$, $Z = 4.66$, $r = 0.26$	S7. $p < .001$, $Z = 5.45$, $r = 1.00$ S12. $p < .001$, $Z = 3.29$, $r = 0.39$ S16. $p < .001$, $Z = 8.13$, $r = 0.49$	**Supported**
H3. TCT of observation would be similar to the N-CPM for complex tasks	S3. $p = .26$, $Z = 0.68$, $r = 0.17$ S4. $p = .54$, $Z = -0.61$, $r = 0.14$ S8. $p = .94$, $Z = -0.07$, $r = 0.02$ S9. $p = .47$, $Z = -0.72$, $r = 0.11$ S10. $p = .11$, $Z = -1.79$, $r = 0.22$ S11. $p = .42$, $Z = -0.80$, $r = 0.21$ S13. $p = .80$, $Z = -0.25$, $r = 0.04$ S14. $p = .36$, $Z = 0.91$, $r = 0.16$ S15. $p = .92$, $Z = -0.02$, $r = 0.00$	S1. $p = .50$, $Z = -1.14$, $r = 0.33$ S7. $p = 1.00$, $Z = 0.00$, $r = 0.00$ S16. $p = .07$, $Z = -1.83$, $r = 0.15$	**Not rejected**
		S12. $p < .001$, Z = −1.82, r = 0.27	**Rejected**

models). Furthermore, it should model the number of errors and provide a remediation strategy.

The second major limitation of N-CPM is its limited capability to quantify driving performance. The ACT-R or QN family of models could estimate lane deviation, steering angle, or acceleration. In future versions of the N-CPM, we plan to enhance the model to estimate novices' driving performance and validate the model with data from novice and expert drivers. In addition, the next version of the N-CPM should include the updated activation function. In ACT-R, activation is the summation of base-level activation (chunks' usefulness to past experiences), spreading activation (chunk's relevance to the current context), and partial matching activation (chunk's similarity to other chunks) (Leiden and Best, 2005). Experts have these three activations, but novices lack past experiences or knowledge. Thus, the activation level of novices can be lower than that of experts which could lead to lower recall probability. However, for this study, implementing this concept into the N-CPM was not possible as the activation function in current version of the N-CPM represents the base-level activation, which cannot directly be compared to the activation function in ACT-R. The third limitation is the model's generalizability. The N-CPM was only validated with a specific population (i.e., novice LEO) in this study. Future studies should validate the model with novices in other domains to improve the model's generalizability.

Lastly, there were some technical and experimental limitations. We used a video analysis software with 30 fps, which could not exactly pinpoint the onset and end of the tasks in millisecond level. Future studies should use a higher resolution analysis software to generate more precise time estimates. Officers' level of experience could also affect the outcomes. Although participants had to have less than five years of experience as a primary patrol officer to be included in the study, they were not exactly equivalent in their amount of practice with each scenario. Furthermore, issues related to lack of normality and equal variance might have been due the naturalistic study settings. For example, officers' duties or time of the day can impact the frequency or duration in which officers used in-vehicle technologies. Some of our observations were conducted in mornings and some were conducted at noon or during afternoon shifts. The differences in the shift and traffic level might cause differences in officers' duties and use of in-vehicle technologies. In addition, driving was always the primary task and the scenarios were secondary tasks for LEOs. The model does not account for the activities required in driving. Therefore, variation within individuals may arise because the LEOs were in different driving conditions as the secondary scenarios arose.

5. Conclusion

The objective of this study was to propose an extension of CPM-GOMS for novice law enforcement officers to model their performance while interacting with in-vehicle technologies. To validate the model, we conducted a ride-along study with nLEOs. The findings suggested that the N-CPM could estimate nLEO's performance for simple and complex tasks. Future studies will focus on applying the N-CPM model to more general driving domain tasks to validate the usefulness of the N-CPM for modeling novice driver performance. The model was developed in R package and released to Github and can be downloaded for free. Future versions of N-CPM should include fundamental human information processing principles and model driving performance similar to methods such as ACT-R or QN family of models.

Declaration of competing interest

None.

Acknowledgement

Funding for this research was provided by the National Science Foundation (No. IIS-2041889). The views and opinions expressed are those of authors. The authors would like to thank Jordan Nunn, Mary Gamble, Kaiwen Wang, Azima Maredia, and Ayodamola Sanni for their help with the data collection and post-processing.

References

Al Seraj, M.S., Pastel, R., Al-Hasan, M., 2018. ASurvey on USER MODELING IN HCI.
Altmann, E.M., Schunn, C.D., 2019. Integrating decay and interference: a new look at an old interaction. In: Paper Presented at the Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society.
Anderson, J.R., Matessa, M., Lebiere, C., 1997. ACT-R: a theory of higher level cognition and its relation to visual attention. Hum. Comput. Interact. 12 (4), 439–462.
Anderson, J.R., Reder, L.M., Lebiere, C., 1996. Working memory: activation limitations on retrieval. Cognit. Psychol. 30 (3), 221–256.
BLS, 2019. Table A-6. Fatal Occupational Injuries Resulting from Transportation Incidents and Homicides by Occupation, All United States, 2015, 2016.
BLS, 2020. Fatal Occupational Injuries to Emergency Responders.
Borowsky, A., Shinar, D., Parmet, Y., 2008. Sign location, sign recognition, and driver expectancies. Transport. Res. F Traffic Psychol. Behav. 11 (6), 459–465.
Bothell, D., 2017. ACT-R 7 Reference Manual. Available at: act-r. psy. cmu. edu/wordpress/wpcontent/themes/ACT-R/actr7/reference-manual. pdf.
Bothell, D., 2020. ACT-R 7.21+ Reference Manual. Retrieved from.
Cao, S., Qin, Y., Jin, X., Zhao, L., Shen, M., 2014. Effect of driving experience on collision avoidance braking: an experimental investigation and computational modelling. Behav. Inf. Technol. 33 (9), 929–940.
Card, M.O.R.A.N., Newell, 1986. The model human processor- an engineering model of human performance. Handbook of perception and human performance 2 (45–1).
Carmichael, A., Larson, A., Gire, E., Loschky, L., Rebello, N.S., 2010. How does visual attention differ between experts and novices on physics problems?. In: Paper Presented at the AIP Conference Proceedings.
Chase, W.G., Simon, H.A., 1973. Perception in chess. Cognit. Psychol. 4 (1), 55–81.
Christie, R., 2001. The effectiveness of driver training as a road safety measure: an international review of the literature. In: Paper Presented at the Road Safety Research, Policing and Education Conference, 2001, Melbourne, Victoria, Australia.
Das, A., Stuerzlinger, W., 2007. A cognitive simulation model for novice text entry on cell phone keypads. In: Paper Presented at the Proceedings of the 14th European Conference on Cognitive Ergonomics: Invent! Explore!.
Davids, K., Bennett, S., Newell, K.M., 2006. Movement system variability. Human kinetics.
Dehban, A., Menhaj, M., Sajedin, A., 2015. Neuro-ACT cognitive architecture applications in modeling driver's steering behavior in turns. AUT Journal of Modeling and Simulation 47 (2), 21–29.
Di Nota, P.M., Huhta, J.-M., 2019. Complex motor learning and police training: applied, cognitive, and clinical perspectives. Front. Psychol. 10 (1797) https://doi.org/10.3389/fpsyg.2019.01797.
Drummond, A.E., 1989. An Overview of Novice Driver Performance Issues: A Literature Review.
Dudzik, K.A.T., 2019. Cognitive Modeling as a Method for Agent Development in Artificial Intelligence. Carleton University.
Eddelbuettel, D., Francois, R., Allaire, J., Ushey, K., Kou, Q., Russell, N., Bates, D., Chambers, J., Eddelbuettel, M.D., 2022. Package 'Rcpp'.
Embrey, D., 2005. Understanding human behaviour and error. Human Reliability Associates 1 (2005), 1–10.
Estes, S., 2015. The workload curve: subjective mental workload. Hum. Factors 57 (7), 1174–1187.
Estes, S., 2017. Cogulator. The MITRE Corporation.
Estes, S., 2021. Cogulator - A Cognitive Modeling Calculator.
Estes, S., Masalonis, A.J., 2003. I see what you're thinking: using cognitive models to represent working memory usage for traffic flow management decision support prototypes. In: Paper Presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
Filtness, A., Mitsopoulos-Rubens, P.E., Lenne, M., 2013. The impact of interface modality on police officers' visual behaviour when using an in-vehicle system. In: Paper Presented at the Proceedings of the 3rd International Conference on Driver Distraction and Inattention (DDI 2013), Gothenburg, Sweden.
Fincham, J.M., 2005. Cognitive Modeling and fMRI: an Integrated Approach toward Understanding Mechanisms of Complex Skill Performance. Carnegie Mellon University.
Fitts, P.M., Posner, M.I., 1967. Human Performance.
Galy, E., Cariou, M., Mélan, C., 2012. What is the relationship between mental workload factors and cognitive load types? Int. J. Psychophysiol. 83 (3), 269–275.
Goode, S.L., Meeuwsen, H.J., Magill, R.A., 1998. Benefits of providing cognitive performance strategies to novice performers learning a complex motor skill. Percept. Mot. Skills 86 (3), 976–978.
Green, P., 1999. The 15-second rule for driver information systems. In: Paper Presented at the Proceedings of the ITS America Ninth Annual Meeting.
Green, P., 2008. Driver Interface/HMI Standards to Minimize Driver Distraction/overload. Retrieved from.
Hembroff, C.C., Arbuthnott, K.D., Krätzig, G.P., 2018. Emergency response driver training: dual-task decrements of dispatch communication. Transport. Res. F Traffic Psychol. Behav. 59, 222–235.
Hick, W.E., 1952. On the rate of gain of information. Q. J. Exp. Psychol. 4 (1), 11–26.
Hyman, R., 1953. Stimulus information as a determinant of reaction time. J. Exp. Psychol. 45 (3), 188.
Jeffrey Bolkhovsky, M., Ritter, F.E., Chon, K., Qin, M., 2018. Performance Trends during Sleep Deprivation.
John, B.E., Gray, W.D., 1995. CPM-GOMS: an analysis method for tasks with parallel activities. In: Paper Presented at the Conference Companion on Human Factors in Computing Systems.
John, B.E., Kieras, D.E., 1996. Using GOMS for user interface design and evaluation: which technique? ACM Trans. Comput. Hum. Interact. 3 (4), 287–319.
Kavakli, M., Gero, J.S., 2003. Strategic Knowledge Differences between an Expert and a Novice Designer *Human Behaviour in Design*. Springer, pp. 42–52.
Kelley, C.R., 1969. What is adaptive training? Hum. Factors 11 (6), 547–556.
Kieras, D., 1997. A Guide to GOMS Model Usability Evaluation Using NGOMSL *Handbook Of Human-Computer Interaction*. Elsevier, pp. 733–766.
Kieras, D.E., 1999. A Guide to GOMS Model Usability Evaluation Using GOMSL and GLEAN3. University of Michigan, p. 313.
Kieras, D.E., Meyer, D.E., 1997. An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. Hum. Comput. Interact. 12 (4), 391–438.
Kim, D., Myung, R., 2016. A unit touch gesture model of performance time prediction for mobile devices. Journal of the Ergonomics Society of Korea 35 (4), 277–291.
Kozhevnikov, M., 2007. Cognitive styles in the context of modern psychology: toward an integrated framework of cognitive style. Psychol. Bull. 133 (3), 464.
Kundel, H.L., Nodine, C.F., 1975. Interpreting chest radiographs without visual search. Radiology 116 (3), 527–532.
Laird, J., Hucka, M., Huffman, S., Rosenbloom, P., 1991. An analysis of Soar as an integrated architecture. ACM SIGART Bull. 2 (4), 98–103.
Laird, J.E., Congdon, C.B., 2015. The Soar User's Manual Version 9.5. 0. The Regents of the University of Michigan.
Law, B., Atkins, M.S., Kirkpatrick, A.E., Lomax, A.J., 2004. Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment. In: Paper Presented at the Proceedings of the 2004 Symposium on Eye Tracking Research & Applications.
Lee, S.C., Yoon, S.H., Ji, Y.G., 2019. Modeling task completion time of in-vehicle information systems while driving with keystroke level modeling. Int. J. Ind. Ergon. 72, 252–260.
Leiden, K., Best, B., 2005. A Cross-Model Comparison of Human Performance Modeling Tools Applied to Aviation Safety, vol. 80301. Micro Analysis & Design, Inc., Boulder, CO, 2005.
Lin, C.-J., Wu, C., 2012. Mathematically modelling the effects of pacing, finger strategies and urgency on numerical typing performance with queuing network model human processor. Ergonomics 55 (10), 1180–1204.
Liu, K., 2019. Measuring and Quantifying Driver Workload on Limited Access Roads.
Liu, Y., Feyen, R., Tsimhoni, O., 2006a. Queueing Network-Model Human Processor (QN-MHP) A computational architecture for multitask performance in human-machine systems. ACM Trans. Comput. Hum. Interact. 13 (1), 37–70.
Liu, Y., Feyen, R., Tsimhoni, O., 2006b. Queueing Network-Model Human Processor (QN-MHP): a computational architecture for multitask performance in human-machine systems. ACM Trans. Comput. Hum. Interact. 13 (1), 37–70.
Maguire, B.J., Hunting, K.L., Smith, G.S., Levick, N.R., 2002. Occupational fatalities in emergency medical services: a hidden crisis. Ann. Emerg. Med. 40 (6), 625–632.
Manes, D., 1997. Prediction of Destination Entry and Retrieval Times Using Keystroke-Level Models. Retrieved from.
McCaskie, A.W., Kenny, D.T., Deshmukh, S., 2011. How can surgical training benefit from theories of skilled motor development, musical skill acquisition and performance psychology? Med. J. Aust. 194 (9), 463–465.

McKenna, F.P., Crick, J., 1994. Hazard Perception in Drivers: A Methodology for Testing and Training. TRL Contractor Report, vol. 313.
Milton, J., Solodkin, A., Hluštík, P., Small, S.L., 2007. The mind of expert motor performance is cool and focused. Neuroimage 35 (2), 804–813.
Mourant, R.R., Rockwell, T.H., 1972. Strategies of visual search by novice and experienced drivers. Hum. Factors 14 (4), 325–335.
NSC, 2022. Preliminary Semiannual Estimates.
Nyström, A., 2018. Gesture-level Model: A Modified Keystroke-Level Model for Tasks on Mobile Touchscreen Devices.
Oyewole, S.A., Farde, A.M., Haight, J.M., Okareh, O.T., 2011. Evaluation of complex and dynamic safety tasks in human learning using the ACT-R and SOAR skill acquisition theories. Comput. Hum. Behav. 27 (5), 1984–1995.
Paelke, G., 1993. Entry of Destinations into Route Guidance Systems: a Human Factors Evaluation. *Final report*. Retrieved from.
Parasuraman, R., Sheridan, T.B., Wickens, C.D., 2008. Situation awareness, mental workload, and trust in automation: viable, empirically supported cognitive engineering constructs. J. Cogn. Eng. Decis. Mak. 2 (2), 140–160.
Park, J., McKenzie, J., Shahini, F., Zahabi, M., 2020. Application of cognitive performance modeling for usability evaluation of emergency medical services in-vehicle technology. In: Paper Presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
Park, J., Zahabi, M., 2021. A novel approach for usability evaluation of mobile applications. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 65 (1), 437–441. Sage CA: Los Angeles, CA: SAGE Publications.
Park, J., Zahabi, M., 2022. A Review of Human Performance Models for Prediction of Driver Behavior and Interactions with In-Vehicle Technology. Human factors, 00187208221132740.
Peck, R.C., 2011. Do driver training programs reduce crashes and traffic violations?—a critical examination of the literature. IATSS Res. 34 (2), 63–71.
Purucker, C., Naujoks, F., Prill, A., Neukum, A., 2017. Evaluating distraction of in-vehicle information systems while driving by predicting total eyes-off-road times with keystroke level modeling. Appl. Ergon. 58, 543–554.
Rasmussen, J., 1983. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. IEEE transactions on systems, man, and cybernetics (3), 257–266.
Reason, J., 1990. Human Error. Cambridge university press.
Redding, R.E., 1992. Cognitive Task Analysis of Prioritization in Air Traffic Control.
Ritter, F.E., Schooler, L.J., 2001. The learning curve. International encyclopedia of the social and behavioral sciences 13, 8602–8605.
Rstudio, 2014. Shiny: Easy Web Applications in R.
Salvucci, D.D., 2005. A multitasking general executive for compound continuous tasks. Cognit. Sci. 29 (3), 457–492.
Salvucci, D.D., Chavez, A.K., Lee, F.J., 2004. Modeling Effects of Age in Complex Tasks: A Case Study in Driving.
Salvucci, D.D., Gray, R., 2004. A two-point visual control model of steering. Perception 33 (10), 1233–1248.
Salvucci, D.D., Macuga, K.L., 2002. Predicting the effects of cellular-phone dialing on driver performance. Cognit. Syst. Res. 3 (1), 95–102.
Salvucci, D.D., Mandalia, H.M., Kuge, N., Yamamura, T., 2007. Lane-change detection using a computational driver model. Hum. Factors 49 (3), 532–542.
Salvucci, D.D., Zuber, M., Beregovaia, E., Markley, D., 2005. Distract-R: rapid prototyping and evaluation of in-vehicle interfaces. In: Paper Presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.
Schneegaß, S., Pfleging, B., Kern, D., Schmidt, A., 2011. Support for modeling interaction with automotive user interfaces. In: Paper Presented at the Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications.
Shahini, F., Zahabi, M., Patranella, B., Mohammed Abdul Razak, A., 2020. Police officer interactions with in-vehicle technologies: an on-road investigation. In: Paper Presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
Sharif, B., Falcone, M., Maletic, J.I., 2012. An eye-tracking study on the role of scan time in finding source code defects. In: Paper Presented at the Proceedings of the Symposium on Eye Tracking Research and Applications.
Shupsky, T., Lyman, A., He, J., Zahabi, M., 2021. Effects of mobile computer terminal configuration and level of driving control on police officers' performance and workload. Hum. Factors 63 (6), 1106–1120.
Smiley, A., Smahel, T., Donderi, D., Caird, J., Chisholm, S., Lockhart, J., Teteris, E., 2007. The Effects of Cellphone and Cd Use on Novice and Experienced Driver Performance.
Sohn, Y.W., Doane, S.M., 2003. Roles of working memory capacity and long-term working memory skill in complex task performance. Mem. Cognit. 31 (3), 458–466.
St Amant, R., Horton, T.E., Ritter, F.E., 2004. Model-based evaluation of cell phone menu interaction. In: Paper Presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.
Tsimhoni, O., Reed, M.P., 2007. The virtual driver: integrating task planning and cognitive simulation with human movement models. SAE Trans. 1525–1531.
Underwood, G., 2007. Visual attention and the transition from novice to advanced driver. Ergonomics 50 (8), 1235–1249.
Vila, B., Kenney, D.J., 2002. The prevalence and potential consequences of police fatigue. NIJ Journal 248, 17–21.
Waring, E., Quinn, M., McNamara, A., Rubia, E.A.d. l., Zhu, H., Lowndes, J., Ellis, S., 2022. Skimr: Compact and Flexible Summaries of Data 2, pp. 1–4.
Wickham, H., 2021. Ellipsis 0.3.2.
Wickham, H., Chang, W., 2016. Devtools: tools to make developing r packages easier. R package version 1, 9000, 0.
Wickham, H., Henry, L., Vaughan, D., 2022. Vctrs : Vector Helpers 0.4.1.
Wiedenbeck, S., 1985. Novice/expert differences in programming skills. Int. J. Man Mach. Stud. 23 (4), 383–390.
Wilson, M.D., Boag, R.J., Strickland, L., 2019. All models are wrong, some are useful, but are they reproducible? Commentary on Lee et al. Computational Brain & Behavior 2 (3–4), 239–241, 2019.
Wozniak, D., Park, J., Nunn, J., Maredia, A., Zahabi, M., 2022. Measuring cognitive workload of novice law enforcement officers in a naturalistic driving study. In: Paper Presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
Wu, C., 2018. The five key questions of human performance modeling. Int. J. Ind. Ergon. 63, 3–6.
Wu, C., Liu, Y., 2008. Queuing network modeling of transcription typing. ACM Trans. Comput. Hum. Interact. 15 (1), 1–45.
Yager, C., Dinakar, S., Sanagaram, M., Ferris, T.K., 2015. Emergency Vehicle Operator On-Board Device Distractions. Texas A&M Transportation Institute Technical Report, 2015, pp. 1–50.
Yang, X., Hyup Kim, J., Nazareth, R., 2019. Hierarchical task analysis for driving under divided attention. In: Paper Presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
Zahabi, M., 2017. Analysis and Redesign of Police Vehicle Mobile Computer Terminal for Minimizing Officer Driving Distraction.
Zahabi, M., Abdul Razak, A.M., 2020. Adaptive virtual reality-based training: a systematic literature review and framework. Virtual Real. 24 (4).
Zahabi, M., Kaber, D., 2018a. Effect of police mobile computer terminal interface design on officer driving distraction. Appl. Ergon. 67, 26–38.
Zahabi, M., Kaber, D., 2018b. Identification of task demands and usability issues in police use of mobile computing terminals. Appl. Ergon. 66, 161–171.
Zahabi, M., Lyman, A., 2019. Impact of electronic medical records on patient-provider communication. In: Paper Presented at the Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care.
Zahabi, M., Pankok Jr., C., Park, J., 2020a. Human factors in police mobile computer terminals: a systematic review and survey of recent literature, guideline formulation, and future research directions. Appl. Ergon. 84, 103041.
Zahabi, M., Park, J., Razak, A.M.A., McDonald, A.D., 2020b. Adaptive driving simulation-based training: framework, status, and needs. Theor. Issues Ergon. Sci. 21 (5), 537–561.
Zahabi, M., Shupsky, T., Lyman, A., 2019a. Analysis of law enforcement mobile computer terminal interface. In: Paper Presented at the IISE Annual Conference, Orlando, FL.
Zahabi, M., White, M.M., Zhang, W., Winslow, A.T., Zhang, F., Huang, H., Kaber, D.B., 2019b. Application of cognitive task performance modeling for assessing usability of transradial prostheses. IEEE Transactions on Human-Machine Systems 49 (4), 381–387.
Zhang, Y., Wu, C., 2017. Learn to integrate mathematical models in human performance modeling. In: Paper Presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
Zhang, Y., Wu, C., Qiao, C., Sadek, A., Hulme, K.F., 2022. A cognitive computational model of driver warning response performance in connected vehicle systems. IEEE Trans. Intell. Transport. Syst. 23 (9), 14790–14805.