Impact of Decision-Support Tools on Novice Workload in VR

Jin Hong Yu, Charles Rowan, Amela Sadagic, Perry McDowell, Ryan Lee

Naval Postgraduate School

Monterey, CA

advance18th@gmail.com, charles.rowan@nps.edu, asadagic@nps.edu, mcdowell@nps.edu, rslee@nps.edu

Jonathan Vogl

US Army Aeromedical Research Laboratory

Fort Novosel, AL

jonathan.f.vogl.civ@health.mil

ABSTRACT

Novices in shipboard damage control environments face high cognitive demands with limited experience, which often lead to critical decision errors. This research investigates whether automated decision support tools (DSTs) can reduce cognitive workload and improve task performance in novices within a virtual reality simulation. Thirty-three participants completed firefighting scenarios under four DST conditions (No DST, Auditory DST, Visual DST, Multimodal DST), while performance, subjective workload ratings, and physiological workload indices (HRV, GSR), were recorded. Results indicated a robust multivariate effect ($p < .001$) and a statistically significant discriminant function based on performance metrics; however no significant differences emerged across cognitive workload indices in either multivariate or univariate analyses. Performance scores were consistently higher in all DST-supported conditions, with the auditory modality showing the strongest benefit (average 10% improvement). These findings suggest that while DSTs may not reliably lower perceived or physiological workload, they nonetheless confer clear operational advantages. Notably, novices appeared largely unaware of the additional cognitive demands imposed by the absence of DSTs, as our second discriminant function based on cognitive workload measures was nonsignificant and did not track the performance decline. This pattern represents a Dunning-Kruger-like effect in which less experienced individuals may lack the metacognitive insight to recognize their own degraded performance or heightened workload. This issue can be further compounded in applied contexts, such as high-fidelity VR simulation, where task demand variability, ecological complexity, and sensory richness can blunt the sensitivity of both subjective and physiological cognitive load indices. This research recommends that the Navy consider implementing auditory DSTs to support novices in visually demanding tasks and guides the future development of decision aids that align the modality of support with the nature of the operator's task.

ABOUT THE AUTHORS

Jin Hong Yu is a Lieutenant Commander in the Republic of Korea Navy and a surface warfare officer serving at Task Fleet Command, Korea. He is a graduate of the United States Naval Academy and the Naval Postgraduate School.

Lieutenant Colonel Charles Rowan, Ph.D., is a US Army Simulation Operations Officer serving as the Interim Director of the Modeling, Virtual Environments, and Simulation (MOVES) Institute at the Naval Postgraduate School in Monterey, CA. He is a graduate of the United States Military Academy and the Naval Postgraduate School.

Amela Sadagic, Ph.D., is a Research Associate Professor and a Co-director of the Center for Advanced Manufacturing and the Consortium for Advanced Manufacturing Research and Education (CAMRE) at the Naval Postgraduate School in Monterey, CA. Dr. Sadagic holds a Ph.D. in Computer Science from University College London, UK.

Perry McDowell is a Faculty Associate for Research at the MOVES Institute at the Naval Postgraduate School in Monterey, CA. Mr. McDowell is a graduate of the United States Naval Academy and the Naval Postgraduate School.

Ryan Lee is a Technical Artist for the Future Tech team at the Naval Postgraduate School in Monterey, CA.

Jon Vogl, Ph.D., is a research psychologist at the US Army Aeromedical Research Laboratory at Fort Novosel, Alabama. Dr. Vogl is a graduate of South Dakota State University and the University of South Dakota.

Impact of Decision-Support Tools on Novice Workload in VR

Jin Hong Yu, Charles Rowan, Amela Sadagic, Perry McDowell, Ryan Lee

Naval Postgraduate School

Monterey, CA

advance18th@gmail.com, charles.rowan@nps.edu, asadagic@nps.edu, mcdowell@nps.edu, rslee@nps.edu

Jonathan Vogl
US Army Aeromedical Research Laboratory
Fort Novosel, AL
jonathan.f.vogl.civ@health.mil

INTRODUCTION

Background

The U.S. military is facing an unprecedented recruiting crisis with a collective shortfall of approximately 41,000 recruits in fiscal year 2023, which is fundamentally challenging force structure and operational readiness (Vergun, 2023). These manpower shortfalls are forcing services to rely more heavily on existing personnel to fill critical roles that exceed their domain-specific expertise, making them novices in areas that they must critically fill. This challenge is particularly acute in shipboard damage control operations, where every crew member must be prepared to respond effectively to emergencies regardless of their occupational specialty. Historical incidents like the USS Forrestal fire in 1967 and the USS Cole bombing in 2000 demonstrate how deficiencies in damage control training and experience can lead to catastrophic outcomes. Furthermore, a comprehensive evaluation by the United States Government Accountability Office (GAO) revealed that increased deployment schedules combined with crew reductions have resulted in insufficient training opportunities and expired certifications for sailors (Pendleton, 2017).

This personnel shortage and insufficient training directly impact operations such as shipboard damage control. Novice operators who are limited in experience and lack well-developed cognitive frameworks can be overwhelmed by high-stress damage control situations, including monitoring multiple systems, assessing environmental conditions, coordinating with team members, and executing complex procedures under extreme time pressure and stress. Research in theories of cognitive workload demonstrates that individuals possess finite mental capacity, and exceeding these limits results in degraded performance, increased error rates, reduced coordination effectiveness, and potential catastrophic outcomes. These challenges are compounded in modern operational environments where technological advancements have increased both the volume and complexity of information available to decision makers.

Recognizing these interconnected challenges of personnel shortages and cognitive overload, the Department of the Navy (DON) has prioritized technological innovation to enhance decision-making capabilities, such as automated decision support tools (DST) (U.S. Navy, 2024). These systems could help novice operators by providing real-time assistance to simplify decision-making processes, reduce information processing burdens, and guide operators through complex procedures during critical damage control operations. However, to be effective, DSTs must align with human cognitive processes and limitations. Systems that fail to account for user cognitive characteristics may inadvertently increase workload, create confusion, or even degrade performance. This challenge is particularly acute for novice operators, who may lack the expertise necessary to integrate automated assistance effectively into their own decision-making processes.

This research effort addresses a critical gap in understanding how automated DSTs affect novice operators in high-stakes operational environments. While extensive research has examined the effectiveness of DST with expert operators within their respective domains of expertise, limited work has focused on novice personnel who are now tasked with making more decisions in military operations. Understanding this relationship is essential for developing effective training systems and operational aids that enhance rather than hinder novice performance.

LITERATURE REVIEW

This section reviews the evolution of DSTs and the theoretical underpinnings of cognitive workload. We outline the methods for assessing cognitive workload, highlight research focused on the differences in cognitive processes between novices and experts, and examine the potential of VR technology to serve as a research and training platform.

Evolution of Decision Support Tools

DSTs have evolved significantly since their initial development in the 1980s, when they were conceptualized as computational aids that could compensate for human cognitive limitations (Woods & Roth, 1988). Early implementations followed a prosthetic model where machines served as independent tools to support human decision-makers. However, that approach often failed to account for the complexities of human-machine interaction and the importance of maintaining human control in decision-making processes (Ehn & Kyng, 1984). The recognition of user acceptance issues and cognitive mismatches in early systems led to a fundamental shift toward joint cognitive systems approaches that emphasized human adaptability and control in complex environments (Woods & Hollnagel, 2006; Kontogiannis & Kossiavelou, 1999). Fischer and Reeves (1991) described this evolution as moving toward "computer-based systems that augment a person's ability to create, reflect, design, decide, and reason" (p. 311), emphasizing the collaborative rather than substitutive nature of effective decision support. Military adoption has followed a similar evolutionary pattern, progressing from simple computational aids to sophisticated systems that support complex operational decision-making. Notable examples include naval applications for supervisory control systems for damage control operations (Downs et al., 2002), knowledge-based decision support systems for damage control officers (Calabrese et al., 2012), and emergency planning tools for flooding and fire incidents (Varela & Soares, 2007).

Theoretical Foundations of Cognitive Workload

While cognitive workload definitions vary across research communities, common themes emerge as relating to the finite nature of mental resources and the relationship between task demands and available cognitive capacity (Saleem et al., 2009; Young et al., 2008; Wickens, 2002). Sweller's Cognitive Load Theory (CLT) provides insights into how instructional design affects performance by distinguishing between intrinsic load related to essential task elements, extraneous load imposed by poor design, and germane load associated with productive learning processes (Sweller, 1988, 2011). As technology evolved and human-computer interaction gained importance, CLT progressed to emphasize user interfaces that minimize extraneous load while maximizing productive cognitive processing (Kosch et al., 2023). Wickens' theoretical frameworks provide additional depth to understanding cognitive resource allocation. The Human Information Processing model conceptualizes decision-making as a sequence of sensory processing, perception, cognition, and response selection, with performance dependent on limited resources at each stage (Wickens et al., 2013). The Multiple Resource Theory (MRT) extends this by proposing that humans have separate resource pools for different sensory modalities (visual versus auditory), processing codes (spatial versus verbal), and response types (manual versus vocal), suggesting that tasks requiring resources from different pools can be performed more effectively than those competing for the same resource pool (Wickens, 2002).

Measurement and Assessment of Cognitive Workload

Accurate assessment of cognitive workload is crucial for understanding how DSTs impact human performance, with researchers prioritizing three key measures for assessment techniques (Longo et al., 2022). Performance-based measures assess workload through observable outputs, including task accuracy, response time, and error rates, operating on the assumption that as mental demands exceed cognitive resources, performance degrades (Proctor & Van Zandt, 2008). However, these measures may not detect workload differences when operators maintain performance through increased effort and can be influenced by factors other than cognitive demand (de Waard, 1996). Subjective measures, notably the NASA Task Load Index, evaluate self-reported workload across six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration (Hart & Staveland, 1988; Rubio et al., 2004). Physiological measures provide objective indicators, including heart rate variability (HRV) and galvanic skin response (GSR) which reflect autonomic nervous system activity and sympathetic activation (Boucsein, 2012; Tao et al., 2019). Research best practices emphasize the use of multiple assessment approaches to comprehensively understand the effects of cognitive workload, as the combination of performance, subjective, and physiological measures can provide convergent evidence while addressing the limitations of individual measurement approaches (Hancock & Matthews, 2019; Ayres et al., 2021).

Novice Versus Expert Cognitive Processing

Expert operators have well-developed mental models that allow them to process information and decide the best course of action without extensively comparing options (Klein, 1997). In contrast, novice operators lack these cognitive frameworks and must rely on more resource-intensive analytical processes, requiring significantly more cognitive effort for basic information processing tasks and leaving fewer resources available for higher-level decision-making (Brand-Gruwel et al., 2005; Schubert et al., 2013). This fundamental difference in cognitive efficiency has important implications for how novices and experts can benefit from decision support technologies.

A critical but often overlooked cognitive characteristic of novices is their limited ability to accurately assess their cognitive workload and performance, known as metacognitive miscalibration (Kruger & Dunning, 1999). The Dunning-Kruger effect suggests that individuals with lower analytical reasoning abilities tend to rely on intuitive responses that are generated quickly but often lead to errors, creating a false sense of correctness and resulting in overestimation of performance (Coutinho et al., 2021; Sanchez & Dunning, 2018). This metacognitive miscalibration has important implications for decision support system design and evaluation, as novice operators may not recognize when they need assistance, may not accurately report their cognitive workload, and may not be aware when their performance is degrading (Parasuraman & Manzey, 2010).

Virtual Reality as a Research and Training Platform

Virtual reality (VR) has emerged as a powerful platform for both research and training applications in military and emergency response domains due to its ability to create immersive, controlled environments that replicate operational demands while maintaining experimental control (Parsons, 2015). The military domain has recognized the value of VR in improving readiness while mitigating risks and costs, with applications spanning flight simulation, battlefield training, and equipment familiarization (Arjun & Sanjay, 2024; Pérez-Ramírez et al., 2018). The Navy's Virtual Environments for Ship and Shore Experimental Learning (VESSEL) damage control trainer exemplifies the capability of VR to simulate realistic fire behavior and emergency response coordination, thereby reinforcing decision-making skills under controlled yet realistic conditions (Hussain & Coleman, 2014).

The advantages of VR technology for cognitive workload research stem from its ability to create standardized experimental conditions while maintaining operational realism, thereby overcoming the limitations of the ecological validity of traditional laboratory-based experiments (Makransky & Peterson, 2021). The integration of physiological monitoring with VR simulation provides powerful capabilities for real-time assessment of heart rate, electrodermal activity, and other cognitive state indicators while participants engage in realistic scenarios (Luong et al., 2020; Kman et al., 2023). Despite limitations including simulator sickness and current fidelity constraints, VR platforms enable researchers to study how humans perform in complex operational environments and how they interact with automation in realistic environments and operational contexts (Chittaro et al., 2014).

RESEARCH QUESTIONS AND HYPOTHESES

As the DON continues to face personnel shortages and increased operational demands, there is growing interest in leveraging automated DSTs to enhance the performance of less experienced personnel in critical operations such as shipboard damage control. While these technologies show promise for augmenting human decision-making capabilities, their effectiveness specifically for novice operators in high-stress emergency scenarios requires rigorous empirical evaluation. Unlike experienced personnel who can rely on well-developed mental models and intuitive decision-making frameworks, novices face unique cognitive challenges that may interact differently with automated support systems. Understanding how different modalities of DSTs affect novice cognitive workload and performance is essential for designing effective training systems and operational aids that enhance rather than hinder emergency response capabilities.

Based on the literature review and identified research gaps concerning novice operators and decision support technologies, three research questions and related hypotheses regarding automated DSTs in VR damage control environments were identified and investigated in our study:

- **Research Question 1:** How do automated decision support tools affect the cognitive workload of novices in a synthetic damage control environment? **Hypothesis 1:** Automated DSTs will significantly reduce cognitive workload for novices in a synthetic damage control environment.
- **Research Question 2:** What differences, if any, exist in novices' cognitive workload and performance between different automated decision support tool modalities? **Hypothesis 2:** There is a significant difference between automated DST modalities.
- **Research Question 3:** How do cognitive resource demands imposed by automated decision support tools influence user performance effectiveness? **Hypothesis 3:** There is a negative correlation between cognitive resource demands and user performance affected by automated DSTs.

METHODOLOGY

Overview and Apparatus

This study employed a controlled experimental approach using virtual reality to investigate the effects of automated DSTs on novice cognitive workload in shipboard damage control scenarios. A custom VR simulation, "DC Ready: Crisis Ops," was developed using Unity and deployed on Meta Quest 3 headsets to create an operationally relevant and immersive environment replicating shipboard fire emergencies. The experimental design employed a within-subjects approach where each participant experienced four distinct DST conditions: No DST (control), Visual DST, Auditory DST, and Multimodal DST (visual and auditory combined). This within-subjects design enabled systematic evaluation of different support tool modalities while controlling for individual differences and improving the precision of condition comparisons. The experiment was approved by the Naval Postgraduate School (NPS) Institutional Review Board (IRB) and executed in the Modeling, Virtual Environments, and Simulation (MOVES) Institute laboratory (Figure 3).

Virtual Environment Design

The VR simulation replicated key compartments of naval vessels, where damage control operations commonly occur, specifically the crew's mess and engine room environments, as depicted in Figure 1. Both environments were carefully designed to ensure consistent cognitive workload across locations, avoiding unintended variations due to spatial layout or accessibility differences. We intentionally constrained participants' navigation by blocking corridors, strategically placing objects, and limiting visibility with smoke, to recreate realistic mobility challenges during an actual shipboard emergency. The virtual environment incorporated user interface principles to enhance usability and minimize unnecessary cognitive workload. The system utilized ray casting for intuitive object interaction and placed reporting menus on participants' virtual wrists for natural access while reducing button interfaces to only essential controls to prevent cognitive strain from inefficient interface design.

Figure 1. Crew's Mess Virtual Environment and Fire Scene

Decision Support Tool Implementation

Four distinct DST conditions were systematically implemented to evaluate varying levels of decision support assistance while maintaining operational realism. The control condition (No DST) required participants to rely entirely on observational skills and procedural memory without external assistance. Visual DST utilized heads-up display (HUD) overlays providing real-time visual guidance with text (as shown in Figure 2), which offered non-intrusive task guidance while potentially risking cognitive tunneling where participants might focus excessively on the HUD rather than the dynamic environment. Auditory DST featured real-time verbal instructions delivered through the VR headset's audio system, preserving visual attention for scene monitoring but potentially introducing auditory information overload by means of forced phonological rehearsal of upcoming task actions. The Multimodal DST condition combined both visual and auditory elements simultaneously, testing whether sensory channel integration enhanced or diminished decision-making effectiveness compared to individual modalities.

Figure 2. Visual DST Offering Task Guidance

Task Design and Scenarios

Participants assumed the role of fire team leaders, i.e., individuals responsible for coordinating and executing critical firefighting tasks based on real-world naval procedures. The scenarios incorporated both procedural tasks, which required participants to execute standard protocols, and time-sensitive tasks, which required them to constantly monitor and respond within a time limit. Key tasks included situational assessment for fire classification and suppression direction, continuous monitoring of fire extinguishing status with dynamic particle reignition, disconnection of electrical power sources, removal of flammable materials, real-time monitoring of air tank levels for four team members, and reporting of injured personnel. This combination of task types provided balanced cognitive workload assessment by testing both attentional demands and procedural adherence, ensuring participants remained consistently engaged while enabling precise workload measurement across different DST conditions.

Participants and Test Procedures

A total of thirty-three volunteers (31 male and 2 female, average age = 32.55) were recruited from NPS students and civilian employees, all serving as novices with no prior extensive shipboard damage control experience beyond initial training programs. This scoping allowed for experimental validity. Each participant completed all four DST conditions in an order determined by Latin Square design to systematically control for sequence effects and learning biases. The experimental session began with informed consent, placement of physiological sensors (Shimmer GSR+ and Polar H10 heart rate monitor), and baseline data collection, followed by comprehensive VR training to ensure equal system familiarity across participants. During each five-minute scenario, the VR system recorded task performance while physiological sensors collected real-time data. The participants completed the NASA-TLX assessment, the simulator sickness questionnaire (SSQ), and a custom questionnaire for each scenario. All participants were given three-minute rests between conditions to allow physiological indicators and any potential simulator sickness symptoms to return to baseline levels. All data streams were synchronized through Lab Streaming Layer software to enable precise temporal alignment of subjective, physiological, and performance metrics, allowing for a comprehensive analysis of cognitive workload. The illustrations of the experimental setup are shown in Figure 3. All participant data collected were de-identified.

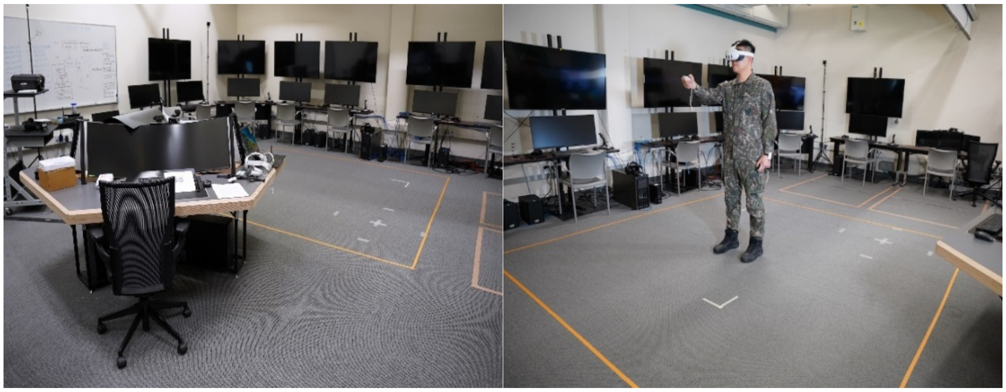

Figure 3. Experiment Setup with VR Headset and Physiological Devices

Variables

The primary independent variable was DST modality, manipulated across four distinct experimental conditions: No DST (control), Visual DST utilizing heads-up display overlays with text, Auditory DST featuring real-time verbal instructions, and Multimodal DST, which combined both visual and auditory elements simultaneously. The dependent variables assessed cognitive workload and task performance through multiple measurement approaches: subjective cognitive workload via NASA-TLX ratings across six dimensions (mental demand, physical demand, temporal demand, performance, effort, and frustration), physiological indicators including HRV measured through inter-beat intervals and GSR measured by the average number of skin conductance response (SCR), and objective performance scores calculated using weighted combinations of task completion rates and error minimization with exponential penalty functions reflecting operational emergency response realities.

Data Collection

The study employed a multi-method approach integrating subjective assessments, objective task performance metrics, and physiological indicators collected in real-time during each scenario. Task performance data were automatically logged by the VR system, tracking task completion rates and error frequencies across all tasks. Physiological data were continuously monitored and recorded using Shimmer GSR+ units for electrodermal activity and Polar H10 chest-worn monitors for HRV. Subjective measures were collected between scenarios using NASA-TLX assessments, simulator sickness questionnaires, and custom difficulty rating scales. All data streams were synchronized through Lab Streaming Layer software to ensure precise temporal alignment across subjective, physiological, and performance metrics, enabling comprehensive analysis of cognitive workload effects under varying DST conditions.

RESULTS

Analysis Approach

A comprehensive statistical analysis was conducted to examine the effects of automated decision support tools on novice cognitive workload and performance in the virtual reality damage control environment. The analysis employed a multivariate approach using repeated-measures multivariate analysis of variance (MANOVA) with an alpha level of 0.05, followed by discriminant function analysis to identify patterns distinguishing the four DST conditions. When significant multivariate effects were detected, univariate repeated-measures analysis of variance (ANOVAs) were performed for each dependent variable, with Holm-Bonferroni correction applied to control for Type I error inflation across multiple comparisons.

The dependent variables included subjective cognitive workload measures (NASA-TLX), physiological indicators (HRV and GSR), and objective performance scores calculated from task completion rates and the number of errors made within the condition. This multi-dimensional approach was essential for answering the research questions about how DSTs affect novice cognitive workload, whether modality differences exist, and how cognitive resource demands influence performance effectiveness.

Multivariate Effects of DST Conditions

The repeated-measures MANOVA revealed a statistically significant multivariate effect of DST conditions on the combined dependent variables (Wilks' Lambda = 0.502, $F(12.0, 246.346) = 6.115$, $p < .001$), confirming that DST type meaningfully influenced novice responses across measured indicators. Discriminant function analysis extracted three functions to distinguish among the four DST conditions, with Function 1 accounting for 97.6% of the variance and showing the strongest canonical correlation ($r = 0.573$). Figure 4's canonical discriminant function plot illustrates the substantial variance captured by Function 1, highlighted by the distinct horizontal dispersion of DST condition centroids along the x-axis. The structure matrix revealed that performance score was the primary discriminating variable (loading = 0.983 on Function 1), while physiological and subjective workload measures loaded primarily on Function 2 contributed minimally to group separation, as evidenced by the limited vertical dispersion in the plot.

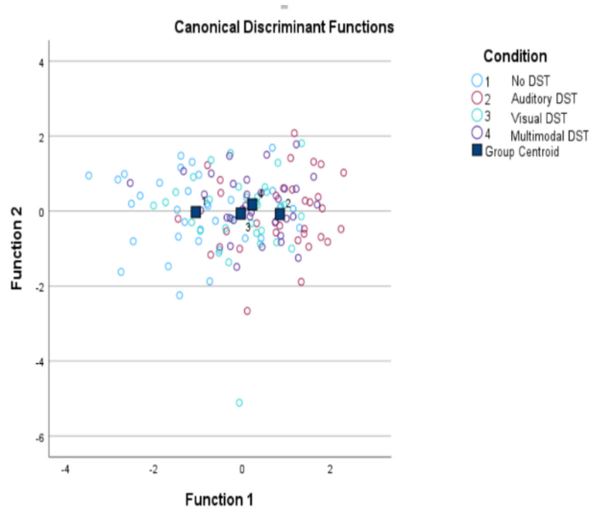

Figure 4. Discriminant Function Plot

Classification accuracy was highest for the No DST condition (69.7%) and Auditory DST condition (66.7%), with substantially lower accuracy for Visual DST (24.2%) and Multimodal DST (15.2%) conditions. This pattern suggests clear behavioral distinctions between the no-support and auditory-support conditions, while visual and multimodal conditions created more variable response patterns among novice operators.

Performance Outcomes and Modality Differences

Univariate analysis revealed significant differences in performance scores across the four DST conditions ($F(2.465, 78.89) = 28.35$, $p < .001$, Greenhouse-Geisser corrected), directly addressing the question of DST's effectiveness for novices. Performance scores showed a clear hierarchy: Auditory DST achieved the highest mean score ($M = 86.64$), followed by Multimodal DST ($M = 82.31$), Visual DST ($M = 80.74$), and No DST ($M = 73.75$). Holm-Bonferroni corrected pairwise comparisons demonstrated that all DST-supported conditions significantly outperformed the No DST baseline (all $p < .001$).

Critically, modality-specific differences emerged among the DST conditions, with Auditory DST significantly outperforming both Visual DST ($p < .001$) and Multimodal DST ($p = .002$). The difference between Visual and Multimodal DST was not statistically significant ($p = .237$). These findings establish both the overall benefit of DST support for novices and the greater effectiveness of auditory delivery in visually demanding damage control scenarios.

Cognitive Workload Measures and Metacognitive Patterns

Despite significant performance improvements with DST support, cognitive workload indicators revealed a striking disconnection between objective performance and subjective experience reported by participants. NASA-TLX subjective workload ratings showed no significant variation across DST conditions ($p = .712$), nor did physiological measures, including heart rate variability (RMSSD, $p = .678$) and galvanic skin response ($p = .722$). This lack of variation in physiological and subjective workload measures occurred despite performance differences of up to 17% between conditions, suggesting that novice participants were unaware of the cognitive resource demands imposed by the absence of decision support. This pattern confirmed that observed workload stability was not due to insufficient measurement sensitivity but rather reflected genuine metacognitive limitations in novice self-assessment capabilities. Simulator sickness questionnaire scores remained stable across DST conditions ($p = .858$) but showed significant cumulative increases across successive experimental runs ($p = .002$), with scores rising from Run 1 ($M = 5.69$) to Run 4 ($M = 72.49$).

Participant Perceptions and Cognitive Resource Utilization

Subjective questionnaire responses reinforced the objective performance hierarchy and provided valuable insights into the competition for cognitive resources. Participants overwhelmingly identified Auditory DST as most helpful (45%, 15 of 33), while No DST (42%) and Visual DST (36%) were most frequently rated as least helpful. Difficulty ratings mirrored performance outcomes, with Auditory DST perceived as easiest ($M = 3.21$) and No DST as most difficult ($M = 2.67$) on a 5-point scale.

Qualitative feedback revealed the underlying mechanism of modality effectiveness, with participants describing visual decision support as "distracting" or causing "information overload," while auditory cues were characterized as "helpful reminders" that supported "task prioritization." These perceptions align with multiple resource theory predictions about visual channel competition in demanding environments. Notably, some participants reported being "too focused on tasks" to process DST information or not recalling visual aids at all, indicating that novices may lack sufficient cognitive resources to effectively utilize competing visual channels during high-demand scenarios.

DISCUSSION

Effects of DSTs on Novice Cognitive Workload

The first research question examined how automated DSTs affect novice cognitive workload in synthetic damage control environments. Results revealed a critical metacognitive disconnect: while DSTs significantly improved performance across all conditions, they produced no measurable reduction in subjective (NASA-TLX) or physiological (HRV, GSR) cognitive workload indicators. This finding contradicts the initial hypothesis that DSTs would reduce cognitive workload but aligns with established research on novice metacognitive limitations and the Dunning-Kruger effect (Coutinho et al., 2021). Novice participants appeared unaware of the additional cognitive demands imposed by the absence of decision support, as evidenced by stable workload measures despite performance deficits of up to 17% in unsupported conditions.

This phenomenon suggests that DSTs provided functional cognitive assistance that was not consciously perceived by novice operators, highlighting the importance of objective performance metrics when evaluating novice responses to technological interventions. The stability of cognitive workload measures across conditions indicates that participants perceived tasks as similarly demanding regardless of actual performance success, demonstrating the overconfidence characteristic of inexperienced individuals. This metacognitive miscalibration has significant implications for training design, as novices may not recognize when they need assistance or accurately assess their own performance limitations. The findings emphasize that while DSTs may not reduce perceived mental effort, they meaningfully enhance operational effectiveness by compensating for cognitive limitations that novices cannot self-identify.

Modality-Specific Effectiveness and Resource Competition

The second research question addressed modality-specific differences among DST types, revealing significant performance variations that confirm the hypothesis of modality-dependent effectiveness. Auditory DST consistently outperformed visual and multimodal conditions, achieving mean performance scores of 86.64 compared to 80.74 and 82.31, respectively, with statistically significant differences confirmed through pairwise comparisons. Participant perceptions reinforced these objective findings, with 45% rating auditory support as most helpful while frequently describing visual aids as "distracting" or causing "information overload." Qualitative feedback revealed that some participants were "too focused on tasks" to process visual DST information, with one participant reporting not recalling visual aids at all during the scenario.

These results align with MRT, which predicts that visual DSTs would compete with already-saturated visual attention channels required for spatial navigation, fire monitoring, and hazard recognition in damage control scenarios (Wickens, 2002). The VR environment demanded continuous visual processing for equipment manipulation, smoke assessment, and spatial orientation, leaving limited visual resources for additional information processing. Conversely, auditory cues accessed underutilized sensory resources, providing decision support without attentional interference or cognitive bottlenecks. The multimodal condition's failure to exceed auditory-only performance suggests that combining modalities may introduce complexity that diminishes the benefits of sensory channel separation, particularly for novice users who lack the cognitive frameworks to efficiently integrate multiple information streams.

Cognitive Resource Demands and Performance Relationships

The third research question explored the relationship between cognitive resource demands and performance effectiveness, revealing a negative correlation between perceived DST complexity and task success that partially supports the hypothesis. Subjective difficulty ratings closely aligned with performance outcomes, with auditory DST perceived as easiest (mean rating 3.21) and yielding the highest performance, while visual DST was rated most difficult (mean rating 2.82) and produced the lowest performance scores. This directional trend indicates that DSTs imposing minimal cognitive burden are most effective for novice users in visually intensive environments, supporting design principles that minimize unnecessary cognitive demand in high-stakes operational scenarios.

However, physiological measures did not differentiate between DST conditions, suggesting that the cognitive burden imposed by different modalities may not have been strong enough to elicit measurable autonomic nervous system responses, or that novices were generally operating near their cognitive capacity regardless of support type. The correlation between subjective effort and performance effectiveness highlights that successful DST design for novices must prioritize intuitive, low-demand interfaces that complement rather than compete with primary task execution. These findings underscore the crucial importance of considering cognitive resource allocation when designing decision support systems for inexperienced operators who may lack the metacognitive awareness necessary to manage competing information streams effectively.

Implications for Naval Operations and Training

These findings have immediate implications for Navy damage control training and operational system design, particularly given ongoing personnel shortages and the push toward decision-making at lower organizational levels outlined in CNO NAVPLAN 2024 (U.S. Navy, 2024). The demonstrated performance improvements with DST support, ranging from 6.9% to 17.4% across conditions, translate to meaningful operational benefits in high-stakes shipboard emergencies where decision accuracy directly impacts sailor safety and vessel survivability. The effectiveness of auditory DSTs suggests that future Navy training systems and operational interfaces should prioritize auditory-based guidance over visual displays in visually demanding environments, potentially reducing the effects of cognitive tunneling while preserving situational awareness critical for damage control effectiveness. Incorporation of auditory DSTs in team intercommunication devices could facilitate shared team understanding and prioritization in damage control scenarios. Further, individual physiological indicators of cognitive workload can be harnessed to dynamically provide DSTs when appropriate and feasible.

The metacognitive disconnect observed in novice participants presents both challenges and opportunities for naval training programs. Since novices cannot accurately assess their own performance limitations, traditional self-directed training approaches may be insufficient for developing damage control competencies. Instead, the Navy should consider integrating DSTs into early-stage training not merely as performance aids, but as compensatory mechanisms for novice metacognitive deficits. This approach could accelerate competency development with multiple repetitions while providing real-time performance feedback to improve self-assessment accuracy, particularly later on when a real-world scenario may not have the benefit of leveraging synthetic DSTs. Furthermore, the findings support the implementation of auditory decision support systems in operational damage control stations, particularly for less experienced crew members who may be called upon during emergencies. Such systems could provide critical task guidance and prioritization cues without interfering with visual monitoring of fire suppression equipment, smoke conditions, and team coordination activities essential for effective damage control operations.

Limitations and Future Research

This study employed a controlled VR simulation, which, although operationally relevant, cannot fully replicate the physical demands, heat exposure, and equipment weight constraints of actual shipboard damage control operations. The moderate difficulty level, designed to ensure novice task completion within experimental timeframes, may have underestimated the cognitive demands and DST benefits that would emerge under more extreme operational stress. Additionally, the exclusive focus on novice participants, while appropriate for addressing current Navy manning challenges, limits comparative insights about DST effectiveness across experience levels.

Future research should investigate adaptive DST systems that dynamically respond to real-time physiological indicators of cognitive overload, potentially optimizing intervention timing and reducing unnecessary cognitive intrusion. Modifying visual DSTs to emphasize and highlight cues with limiting the real estate needed would drive further analysis into multi-modal approaches to delivering DSTs. Expanding the participant pool to include individuals with varying expertise levels would enable an examination of how DST reliance and effectiveness change with experience development. Additionally, incorporating team-based scenarios with communication demands would better approximate the collaborative nature of actual damage control operations and assess whether DST benefits persist under realistic coordination requirements.

Broader Human Performance Implications

The observed metacognitive limitations in novice operators extend beyond naval applications to broader military training and civilian emergency response domains. The finding that performance improvements occurred without conscious awareness of cognitive assistance suggests that objective performance metrics should be prioritized over subjective assessments when evaluating the effectiveness of training for inexperienced personnel. This has implications for designing assessment protocols across military occupational specialties where novices must rapidly develop competency in high-stakes environments.

The modality-specific effectiveness of auditory DSTs also informs human-systems integration principles for complex operational environments. As military systems increasingly incorporate automated decision aids, understanding sensory channel competition becomes critical for optimizing human-machine teaming. The higher performance of auditory guidance in visually demanding tasks suggests design principles that could enhance effectiveness across domains, including aviation, ground combat, and command center activities where visual attention is heavily taxed.

ACKNOWLEDGEMENTS

Sections of this article were previously published as a thesis at NPS [Yu, 2025]. The author would like to acknowledge the support from the Navy Modeling & Simulation Office and the faculty of the NPS, who supported this research. This research is Distribution A, approved for public release; distribution is unlimited. The views expressed in this writing are those of the author and do not reflect the official policy or position of the Department of Defense, U.S. Government, or the Republic of Korea Government.

REFERENCES

Arjun, B. C., & Sanjay, K. R. (2024). Enhancing military training through VR applications. *International Scientific Journal of Engineering and Management*, *3*(5), 1–6. https://doi.org/10.55041/isjem01739

Ayres, P., Lee, J. Y., Paas, F., & van Merriënboer, J. J. G. (2021). The validity of physiological measures to identify differences in intrinsic cognitive load. *Frontiers in Psychology*, *12*, 702538. https://doi.org/10.3389/fpsyg.2021.702538

Boucsein, W. (2012). *Electrodermal activity* (2nd ed.). Springer. https://doi.org/10.1007/978-1-4614-1126-0

Brand-Gruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving by experts and novices: Analysis of a complex cognitive skill. *Computers in Human Behavior*, *21*(3), 487–508. https://doi.org/10.1016/j.chb.2004.10.005

Calabrese, F., Corallo, A., Margherita, A., & Zizzari, A. A. (2012). A knowledge-based decision support system for shipboard damage control. *Expert Systems with Applications*, *39*(9), 8204–8211. https://doi.org/10.1016/j.eswa.2012.01.146

Chief of Naval Operations. (2024). *Navigation Plan 2024*. United States Navy. https://www.navy.mil/Portals/1/CNO/NAVPLAN2024/Files/CNO-NAVPLAN-2024-high-res-v2.pdf

Chittaro, L., Buttussi, F., & Zangrando, N. (2014). Desktop virtual reality for emergency preparedness: User evaluation of an aircraft ditching experience under different fear arousal conditions. *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*, 141–150. https://doi.org/10.1145/2671015.2671025

Coutinho, M. V. C., Thomas, J., Alsuwaidi, A. S. M., & Couchman, J. J. (2021). Dunning-Kruger effect: Intuitive errors predict overconfidence on the cognitive reflection test. *Frontiers in Psychology*, *12*, 1–10. https://doi.org/10.3389/fpsyg.2021.603225

de Waard, D. (1996). *The measurement of drivers' mental workload*. [Thesis fully internal (DIV), University of Groningen]. s.n.

Downs, R., Farley, J. P., & Williams, F. W. (2002). *Damage control automation for reduced manning (DC-ARM) supervisory control system software summary*. Naval Research Laboratory. https://apps.dtic.mil/sti/pdfs/ADA400625.pdf

Ehn, P., & Kyng, M. (1984). *A tool perspective on design of interactive computer support for skilled workers*. [Unpublished manuscript, Swedish Center for Working Life, Stockholm].

Fischer, G., & Reeves, B. (1991). Beyond intelligent interfaces: Exploring, analyzing, and creating success models of cooperative problem solving. *Journal of Applied Intelligence*, *1*, 311–332.

Hancock, P. A., & Matthews, G. (2019). Workload and performance: Associations, insensitivities, and dissociations. *Human factors*, *61*(3), 374-392.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9

Hussain, T. S., & Coleman, S. L. (2014). *Design and development of training games: Practical guidelines from a multidisciplinary perspective*. Cambridge University Press.

Klein, G. (1997). The recognition-primed decision (RPD) model: Looking back, looking forward. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 285–292). Lawrence Erlbaum Associates.

Kman, N. E., Price, A., Berezina-Blackburn, V., Patterson, J., Maicher, K., Way, D. P., McGrath, J., Panchal, A. R., Luu, K., Oliszewski, A., Swearingen, S., & Danforth, D. (2023). First responder virtual reality simulator to train and assess emergency personnel for mass casualty response. *JACEP Open*, *4*(1), e12903. https://doi.org/10.1002/emp2.12903

Kontogiannis, T., & Kossiavelou, Z. (1999). Stress and team performance: Principles and challenges for intelligent decision aids. *Safety Science*, *33*(3), 103–128. https://doi.org/10.1016/S0925-7535(99)00027-2

Kosch, T., Karolus, J., Zagermann, J., Reiterer, H., Schmidt, A., & Woźniak, P. W. (2023). A survey on measuring cognitive workload in human-computer interaction. *ACM Computing Surveys*, *55*(13s), 1–39.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. https://doi.org/10.1037//0022-3514.77.6.1121

Longo, L., Wickens, C. D., Hancock, G., & Hancock, P. A. (2022). Human mental workload: A survey and a novel inclusive definition. *Frontiers in Psychology*, *13*, 883321. https://doi.org/10.3389/fpsyg.2022.883321

Luong, T., Argelaguet, F., Martin, N., & Lecuyer, A. (2020). Introducing mental workload assessment for the design of virtual reality training scenarios. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 662–671. https://doi.org/10.1109/VR46266.2020.00089

Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review*, *33*(3), 937–958. https://doi.org/10.1007/s10648-020-09586-2

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *52*(3), 381–410.

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00660

Pendleton, J. H. (2017). *Navy readiness: Actions needed to address persistent maintenance, training, and other challenges facing the fleet*. United States Government Accountability Office. https://www.gao.gov/products/gao-17-798t

Pérez-Ramírez, M., Zayas-Pérez, B. E., Hernández-Aguilar, J. A., & Ontiveros-Hernández, N. J. (2018). Analytics for military training in virtual reality environments. *Military Applications of Data Analytics*.

Proctor, R.W., & Zandt, T.V. (2008). *Human factors in simple and complex systems* (2nd ed.). CRC Press.

Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology*, *53*(1), 61–86.

Saleem, J. J., Russ, A. L., Sanderson, P., Johnson, T. R., Zhang, J., & Sittig, D. F. (2009). Current challenges and opportunities for better integration of human factors research with development of clinical information systems. *Yearbook of Medical Informatics*, 48–58.

Sanchez, C., & Dunning, D. (2018). Overconfidence among beginners: Is a little learning a dangerous thing? *Journal of Personality and Social Psychology*, *114*(1), 10–28. https://doi.org/10.1037/pspa0000102

Secretary of the Navy. (2024). *Naval science and technology strategy*. United States Navy. https://www.onr.navy.mil/our-research/strategy

Schubert, C. C., Denmark, T. K., Crandall, B., Grome, A., & Pappas, J. (2013). Characterizing novice-expert differences in macrocognition: An exploratory study of cognitive work in the emergency department. *Annals of Emergency Medicine*, *61*(1), 96–109. https://doi.org/10.1016/j.annemergmed.2012.08.034

Sweller, J. (1988) Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285.

Sweller, J. (2011). Cognitive load theory. In K. D. Federmeier (Ed.), *Psychology of Learning and Motivation* (Vol. 55, pp. 37–76). Elsevier. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Tao, D., Tan, H., Wang, H., Zhang, X., Qu, X., & Zhang, T. (2019). A systematic review of physiological measures of mental workload. *International Journal of Environmental Research and Public Health*, *16*(15), 2716.

Varela, J. M., & Soares, C. G. (2007). A virtual environment for decision support in ship damage control. *IEEE Computer Graphics and Applications*, *27*(4), 58–69. https://doi.org/10.1109/MCG.2007.74

Vergun, D. (2023, December 13). *DoD addresses recruiting shortfall challenges*. U.S. Department of Defense. https://www.defense.gov/News/News-Stories/Article/article/3616786/

Wickens, C. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science*, *3*, 159–177. https://doi.org/10.1080/14639220210123806

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering psychology and human performance* (4th ed.). Psychology Press. https://doi.org/10.4324/9781315665177

Woods, D. D., & Hollnagel, E. (2006). *Joint Cognitive Systems*. https://doi.org/10.1201/9781420005684

Woods, D. D., & Roth, E. M. (1988). Cognitive engineering: Human problem solving with tools. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *30*(4), 415–430.

Young, G., Zavelina, L., & Hooper, V. (2008). Assessment of workload using NASA Task Load Index in perianesthesia nursing. *Journal if Perianesthesia Nursing*, *23*(2), 102–110.

Yu, J. H. (2025). *Understanding the effects of automated decision support tools on novice cognitive workload in VR damage control scenarios.* [Thesis, Naval Postgraduate School].