

Mapping Cognitive Fidelity in Joint Cognitive Systems: Neuroergonomics in Simulation-Based Training

Jessica Johnson and Ashley Buczkowski

Old Dominion University, Suffolk, Virginia, VA 23435, USA

ABSTRACT

Assessing whether simulation-based training environments elicit the cognitive demands of real work remains a persistent challenge in human factors and neuroergonomics. This paper presents a meta-analysis of eye-tracking studies (2005–2025) conducted in live, virtual, and constructive simulation-based training environments across aviation, maritime, medical, military, and industrial domains, examining whether inconsistencies in prior findings are explained by variation in cognitive fidelity, which this paper defines as the degree to which simulations preserve the information-processing structure of operational tasks. Across 26 studies cognitive fidelity strongly moderated gaze–performance relationships. High cognitive fidelity simulations produced moderate-to-large and stable effects (mean $r \approx .48$), whereas medium fidelity simulations showed attenuated effects (mean $r \approx .18$) and low fidelity simulations yielded weak and heterogeneous effects (mean $r \approx .07$), independent of simulator realism or eye-tracker resolution. Risks emerged when eye tracking was applied to cognitively shallow or underspecified tasks, where gaze patterns reflected engagement or design artifacts rather than task-relevant cognition. The findings reposition eye tracking as a neuroergonomic diagnostic of cognitive fidelity, yielding actionable guidance for researchers and designers: cognitive work must be engineered before measurement, and eye tracking should be deployed after core design decisions stabilize to evaluate, compare, and refine cognitively faithful training systems.

Keywords: Neuroergonomics, Cognitive fidelity, Eye tracking, Simulation-based training

INTRODUCTION

Simulation-based training is widely used to prepare individuals and teams for complex, high-consequence work across aviation, maritime operations, healthcare, military command-and-control, and industrial safety. These environments enable exposure to rare, dynamic, or hazardous conditions without operational risk (Salas et al., 2009; Alessi, 2000). Historically, training system design has emphasized fidelity, often equated with perceptual or physical realism. However, extensive empirical evidence shows that increased realism alone does not reliably improve learning, transfer, or adaptive performance (Hays & Singer, 1989; Norman et al., 2012; Hamstra et al., 2014).

Training effectiveness depends less on visual resemblance and more on whether simulations reproduce the cognitive work of operational systems,

including how information is distributed, how attention is allocated over time, how uncertainty is resolved, and how coordination unfolds between humans and automation (Vicente, 1999; Hollnagel & Woods, 2005). These properties are referred to as cognitive fidelity, yet they remain inconsistently defined and rarely measured directly, contributing to mixed findings regarding fidelity and training outcomes (Alexander et al., 2005; Hochmitz & Yuviler-Gavish, 2011).

This paper synthesizes 20 years of eye-tracking research in live, virtual, and constructive simulation-based training environments to examine how cognitive fidelity has been instantiated in practice and when gaze-based measures yield interpretable results. Rather than treating eye tracking as a generic measure of attention, we evaluate it as a neuroergonomic signal whose validity depends on the cognitive fidelity of the simulation, providing evidence-based guidance for designing and assessing cognitively faithful training systems.

Cognitive Fidelity and the Limits of Traditional Fidelity Constructs

Classical frameworks distinguish physical, functional, and psychological fidelity (Hays & Singer, 1989), yet applied research has frequently collapsed fidelity into a single dimension of simulator realism. Meta-analytic and experimental studies show that such approaches yield inconsistent results: visually rich simulations may increase extraneous cognitive load or obscure task structure, whereas simpler environments can support robust learning when cognitive demands are well aligned (van Merriënboer & Sweller, 2005; Norman et al., 2012).

Cognitive fidelity differs fundamentally from physical fidelity. It refers to the extent to which a simulation preserves the information-processing structure of the task, including representational constraints, temporal dependencies, and coordination requirements (Vicente, 1999; Hollnagel & Woods, 2005). As a result, a simulation may appear realistic yet lack cognitive fidelity if uncertainty, decision points, or attentional demands are simplified in ways that alter how cognitive work is performed (Hochmitz & Yuviler-Gavish, 2011; Salas et al., 2012). A persistent challenge is that cognitive fidelity is often inferred indirectly through aggregate performance or workload measures, which provide limited insight into moment- to-moment attentional strategies and can mask differences in cognitive execution (Endsley, 1995; Ericsson & Pool, 2016).

Eye Tracking as Neuroergonomic Signal of Cognitive Fidelity

From a neuroergonomic perspective, eye tracking offers a principled means of addressing this measurement gap. Neuroergonomics examines how cognitive processes are shaped by task demands and system design in real-world work contexts (Parasuraman & Rizzo, 2007). Because visual attention is a fundamental cognitive bottleneck, eye movements provide a behavioral window into how attention is allocated in response to task structure, system representations, and environmental uncertainty (Wickens, 2008).

Eye tracking captures not only where operators look, but how attention unfolds over time through fixations, saccades, and related oculomotor dynamics that are sensitive to task demands and cognitive state (Rayner, 1998; Holmqvist et al., 2011). Recent work by Qian et al. (2025) demonstrates that eye-movement signals, while spatially low-dimensional, are rich in temporal structure and encode task-specific and state-specific variance alongside individual differences and measurement noise. From a cognitive fidelity standpoint, this distinction is critical: cognitively faithful simulations should elicit systematic, task-relevant variance in gaze behavior. When such structure is absent, eye tracking reveals a lack of cognitive fidelity rather than a failure of the measurement itself.

When Eye Tracking Helps and Hinders

The literature synthesized in this meta-analysis shows that eye tracking produces interpretable and reproducible effects only when cognitive fidelity is present. In simulations that preserve distributed information, temporal coupling, and meaningful uncertainty, gaze behavior differentiates expertise, reflects anticipatory monitoring, and scales with task demands (Endsley, 2000; Wickens, 2008). In contrast, cognitively shallow simulations, often visually rich but structurally simple, yield gaze patterns dominated by noise or engagement rather than task-relevant cognition.

As Qian et al. (2025) highlight, eye-movement signals conflate multiple sources of variance unless grounded in well-specified task contexts (Holmqvist et al., 2011). Without such grounding, eye tracking can hinder inference by encouraging misattribution of gaze patterns to learner deficiencies rather than to design limitations. From a neuroergonomic perspective, this reflects misalignment between elicited oculomotor dynamics and the cognitive phenomena of interest, not a limitation of eye tracking as a modality.

METHOD

This meta-analysis followed PRISMA 2020 guidelines (Page et al., 2021), with a documented protocol specifying aims, eligibility criteria, coding dimensions, and synthesis strategy to promote transparency and reduce bias. Owing to substantial heterogeneity in eye-tracking measures, simulation modalities, and outcomes, the study was conducted as a theory-driven meta-synthesis rather than a *de novo* pooled-effect meta-analysis, consistent with prior neuroergonomics and human factors reviews of complex sociotechnical systems. The objective was to examine how cognitive fidelity has been instantiated in eye-tracking-based simulation research and to identify conditions under which gaze measures yield interpretable, task-relevant results.

A comprehensive search covered studies published between January 2005 and August 2025, reflecting the adoption of mobile and head-mounted eye tracking in applied training contexts. Searches were conducted using combinations of terms related to eye tracking, simulation-based training, and cognitive outcomes. Reference lists of relevant reviews and foundational cognitive engineering and neuroergonomics sources were also hand-searched.

Studies were included if they: (1) used eye tracking in live, virtual, or constructive simulation-based training or assessment; (2) involved operationally relevant cognitive work (e.g., monitoring, decision making, uncertainty management, coordination); (3) examined outcomes related to performance, situation awareness, workload, error, or learning; and (4) were peer-reviewed or verifiable grey literature. Studies focused on static visual tasks, basic perception, marketing or usability without operational demands, or simulations lacking a training or assessment purpose were excluded. Only English-language publications were considered.

Study Selection

The search process initially identified 1,300 records across all databases. After removing duplicates, 820 titles and abstracts were screened for relevance. Of these, 560 were excluded because they did not meet the basic eligibility criteria, such as lacking a simulation-based task or using eye tracking solely for interface usability without cognitive outcomes. The remaining 240 articles were retrieved for full-text review.

The study selection proceeded in two stages. First, titles and abstracts were independently screened by two reviewers against the predefined inclusion criteria. In the second stage, full-text articles of potentially relevant studies were examined in detail by the same reviewers. Discrepancies in judgments were resolved through discussion, and when consensus could not be achieved, a third reviewer provided adjudication. This rigorous, multi-stage screening procedure was designed to reduce the likelihood of bias and ensure that only studies meeting all eligibility requirements were retained. Ultimately 26 studies were deemed suitable for the quantitative meta-analysis

Data Extraction and Coding

All included studies underwent a structured data extraction process using a predefined coding framework. Study characteristics included domain (e.g., aviation, maritime, healthcare, military, industrial safety), participant population (novices, trainees, professionals), and simulation modality (live, virtual, constructive, or hybrid). To operationalize cognitive fidelity, each study was coded for task and system design features reflecting the preservation of cognitive work, including a) distribution of information across multiple interdependent sources, b) temporal coupling between cues, decisions, and actions, c) presence of uncertainty and need for anticipatory monitoring, and d) degree of human–automation interaction or coordination.

Based on these features, simulations were categorized as exhibiting high, medium, or low cognitive fidelity, following principles from cognitive work analysis and joint cognitive systems theory. Eye-tracking measures were coded by type (e.g., fixation duration, dwell time, scan path structure, transition patterns), level of analysis (global vs. task-relevant areas of interest), and temporal resolution. Studies were also coded for whether gaze data were interpreted in relation to task events or system state, a prerequisite for neuroergonomic validity. Outcome variables were coded by construct (e.g., situation awareness, workload, performance, error) and by whether analyses

examined individual or team-level behavior. For team studies, the presence of coordinated gaze or shared visual attention measures was noted.

Evidence Synthesis

Given heterogeneity in eye-tracking metrics, simulation designs, and outcome measures, the synthesis followed an umbrella meta-synthesis approach rather than recomputing pooled effect sizes across all studies. Quantitative synthesis focused on extracting and summarizing effect sizes from studies with sufficient statistical reporting, emphasizing patterns of association between gaze measures and task-relevant outcomes under different levels of cognitive fidelity. Where multiple gaze metrics were reported within a study, effects were treated as dependent and interpreted conservatively. Effect sizes were summarized as ranges rather than single aggregate estimates to avoid overgeneralization.

A qualitative thematic synthesis was conducted to identify recurring design conditions under which eye tracking produced interpretable versus unstable results. Particular attention was paid to the alignment between gaze measures and cognitive task structure, the role of system context in interpreting eye movements, and common failure modes where eye tracking hindered inference. These quantitative and qualitative strands were integrated to derive neuroergonomics-relevant implications for the use of eye tracking in quantifying, evaluating, and supporting cognitive fidelity in simulation-based training environments.

RESULTS

How Cognitive Fidelity is Instantiated in Practice

The results demonstrate that eye tracking yields interpretable, reproducible insights only when simulations preserve the cognitive structure of real work. Cognitive fidelity emerged as a dominant moderator of gaze-performance relationships, accounting for both the magnitude and heterogeneity of observed effects. High cognitive fidelity simulations produced stronger, more stable effects with narrower confidence intervals (mean $r = .48$), whereas low fidelity simulations produced weak, heterogeneous results (mean $r = .07$) regardless of technological sophistication. These findings position eye tracking as a neuroergonomic diagnostic of cognitive fidelity, rather than as a general-purpose measure of attention or engagement opined in prior research literature (Parasuraman & Rizzo, 2007; Wickens, 2008). Across the studies, cognitive fidelity was rarely defined explicitly. Instead, it was instantiated implicitly through combinations of task design, information representation, and system coupling. Examination of these design features revealed three empirically distinguishable categories reflecting high, medium, and low cognitive fidelity, consistent with the view that fidelity relevant to cognition is a property of task-representation-interaction systems rather than surface realism (Vicente, 1999; Hollnagel & Woods, 2005; Hamstra et al., 2014).

High cognitive fidelity simulations were characterized by distributed information across multiple interdependent sources, strong temporal coupling

between cues, decisions, and actions, the presence of uncertainty requiring anticipatory monitoring, and meaningful coordination between humans and automation. These characteristics align with cognitive systems engineering accounts of complex work in which cognition is shaped by constraints and representations embedded in the system (Vicente, 1999; Hollnagel & Woods, 2005). In the reviewed corpus, high cognitive fidelity designs were most prevalent in full-mission maritime bridge simulators, aviation simulators, and scenario-driven medical simulations.

Medium cognitive fidelity simulations preserved task-relevant information but simplified temporal dynamics or reduced coordination demands. These environments commonly emphasized procedural execution and constrained variability, which can reduce opportunities for trainees to engage in adaptive sensemaking and attention management (Hochmitz & Yuviler-Gavish, 2011; Salas et al., 2012). As a result, the range of attentional strategies expressed in gaze data was typically narrower.

Low cognitive fidelity simulations relied on static displays, scripted task flows, minimal uncertainty, or single-focus interfaces. Such designs are consistent with the broader fidelity literature showing that perceptual complexity can be decoupled from the cognitive structure required for learning and transfer (Norman et al., 2012; Hamstra et al., 2014). These designs were common in early XR training demonstrations and proof-of-concept systems where task demands were simplified. Importantly, over 80% of studies eligible for quantitative synthesis fell within the high cognitive fidelity category, indicating that simulations capable of supporting interpretable gaze–performance relationships were disproportionately those that preserved the structure of real cognitive work (Vicente, 1999; Hollnagel & Woods, 2005).

Quantitative Relationships Between Gaze and Task Performance

Among the 26 studies reporting sufficient statistical detail for quantitative extraction, gaze–performance relationships were strongly moderated by cognitive fidelity. Across domains, consistent and interpretable effects emerged primarily in simulations classified as high in cognitive fidelity, supporting the broader neuroergonomics premise that attentional allocation is shaped by task demands and system design constraints (Parasuraman & Rizzo, 2007; Wickens, 2008). In high cognitive fidelity simulations, gaze measures anchored to task-relevant information sources reliably differentiated expertise and performance. Reported effect sizes clustered in the moderate-to-large range, with correlations typically between $r \approx .40$ and $.55$ and standardized mean differences between $g \approx 0.6$ and 0.8 . These effects were most consistent when gaze metrics reflected functionally meaningful behaviors such as anticipatory monitoring, efficient sampling of predictive cues, and structured transitions among interdependent information sources. This pattern aligns with theory and evidence showing that expert performance is supported by selective attention to diagnostically relevant cues and effective situation awareness under dynamic conditions (Endsley, 1995, 2000; Wickens, 2008).

By contrast, medium cognitive fidelity simulations exhibited attenuated and more variable effects. Gaze–performance associations were generally

small-to- moderate ($r \approx .10-.25$), consistent with reduced demands for adaptive monitoring and uncertainty management. This attenuation is compatible with training research demonstrating that when tasks are simplified or decision points are constrained, differences in cognitive strategy become less observable and less consequential for performance outcomes (Hochmitz & Yuviler-Gavish, 2011; Norman et al., 2012).

In low cognitive fidelity simulations, gaze–performance relationships were weak or near zero ($r \approx .00-.15$). Studies in this category frequently reported null findings or highly variable results even when using high-quality sensors. This pattern is consistent with the broader measurement principle that eye movements are informative only when the task elicits structured, goal-directed attentional behavior rather than exploratory viewing (Rayner, 1998; Holmqvist et al., 2011). It also aligns with fidelity research showing that realism does not guarantee transfer when cognitive demands are not preserved (Norman et al., 2012; Hamstra et al., 2014). Overall, these patterns indicate that gaze–performance relationships scale with cognitive fidelity, not with simulator realism or eye-tracker resolution (Vicente, 1999; Hamstra et al., 2014).

Effect Size Differences by Cognitive Fidelity

To examine heterogeneity systematically, a moderator analysis was conducted with cognitive fidelity level as the primary moderator. Cognitive fidelity functioned as a strong moderator of gaze–performance relationships, with higher fidelity simulations yielding both stronger effects and reduced heterogeneity. As shown in Figure 1, high cognitive fidelity simulations produced the largest and most stable effects, with a mean correlation of $r = .48$ and a relatively narrow confidence interval ($\approx [.40, .55]$). This indicates both substantively meaningful effects and reduced variability when gaze measures are aligned with cognitively faithful task designs. From a cognitive engineering perspective, this stability is expected because high-fidelity task structures constrain attention in consistent, interpretable ways (Vicente, 1999; Hollnagel & Woods, 2005).

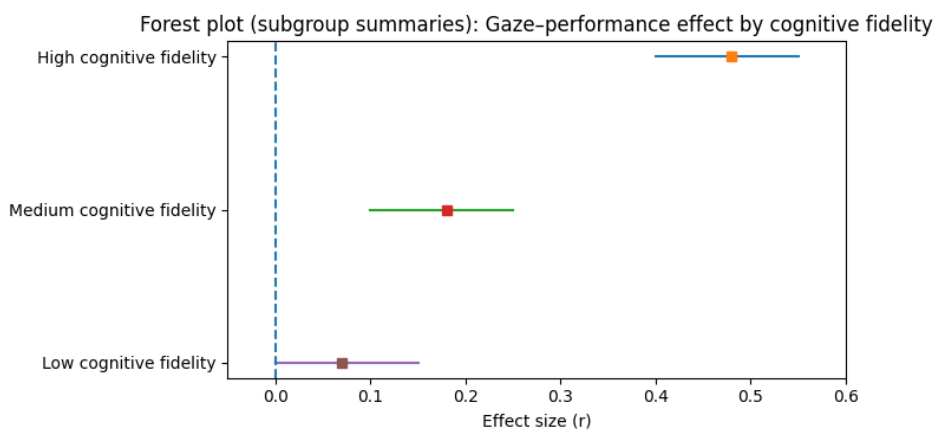


Figure 1: Forest plot of subgroup summary effects (correlations, r) relating gaze-based measures to task performance, stratified by cognitive fidelity level.

Figure 1, above, presents a forest plot of subgroup summary effects, illustrating both the magnitude and heterogeneity of gaze–performance relationships as a function of cognitive fidelity. High cognitive fidelity simulations yielded the largest mean effect ($r = .48$) and the narrowest confidence interval ($\approx [.40, .55]$), indicating both substantive effects and reduced variability across studies. Medium fidelity simulations showed attenuated effects ($r = .18$) with wider confidence intervals, while low fidelity simulations exhibited near-zero effects ($r = .07$) and the greatest uncertainty. Importantly, the progressive widening of confidence intervals from high to low cognitive fidelity indicates that effect size heterogeneity was structured rather than random. Variability increased systematically as simulations departed from cognitively faithful task designs, suggesting that differences in task structure, not measurement noise or domain, were the primary drivers of heterogeneity.

In contrast, medium cognitive fidelity simulations yielded a mean effect size of $r = .18$ with wider confidence intervals ($\approx [.10, .25]$), reflecting increased variability across study designs and gaze operationalizations. Low cognitive fidelity simulations exhibited the smallest effects, with a mean correlation of $r = .07$ and confidence intervals spanning near zero to small positive values ($\approx [.00, .15]$). The widening of confidence intervals across fidelity levels is consistent with eye-tracking measurement theory: when task demands do not impose strong structure on gaze behavior, variance attributable to individual differences and measurement pipeline factors becomes more prominent (Holmqvist et al., 2011; Qian et al., 2025). Critically, the moderator pattern indicates that effect size heterogeneity was structured rather than random. Variability decreased systematically as cognitive fidelity increased, supporting the interpretation that heterogeneity was driven primarily by differences in task design and representational structure rather than by sensor limitations (Hamstra et al., 2014; Qian et al., 2025).

To examine whether cognitive fidelity systematically moderated the relationship between gaze-based measures and task performance, a subgroup moderator analysis was conducted, see Figure 2 below. Studies were grouped according to high, medium, or low cognitive fidelity based on task design features identified during qualitative coding. Mean effect sizes and confidence intervals were computed for each subgroup to assess both the magnitude and stability of gaze–performance relationships across fidelity levels.

These moderator patterns suggest that weak or inconsistent gaze effects are not failures of eye tracking itself, but consequences of insufficient cognitive fidelity—an interpretation examined further by analyzing recurrent failure modes in the literature.

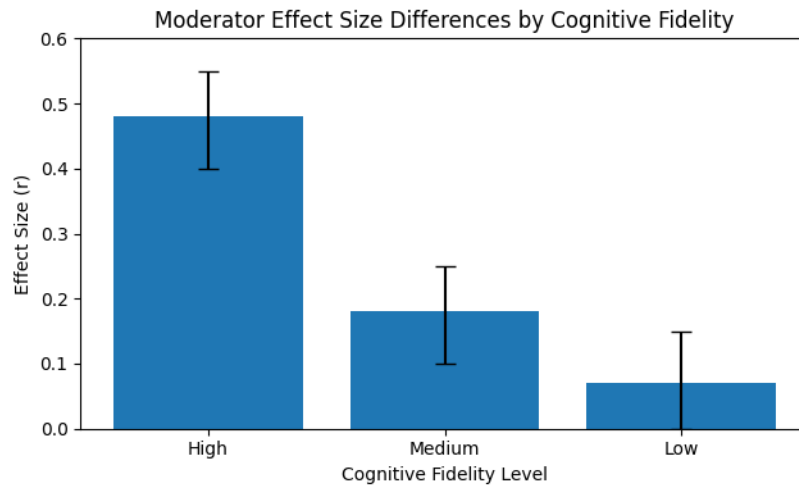


Figure 2: Mean gaze-performance effect sizes (r) by cognitive fidelity level. Error bars represent confidence intervals. High cognitive fidelity simulations yield substantially larger and more stable effects than medium and low fidelity simulations, indicating that cognitive fidelity moderates the interpretability of eye-tracking measures.

Conditions Under Which Eye Tracking Fails to Produce Interpretable Results

Three recurrent failure modes were identified in studies where eye tracking failed to yield interpretable or reliable results: 1) underspecified cognitive tasks, 2) cognitively shallow XR environments, and 3) lack of system context. These failure modes were closely associated with low or insufficient cognitive fidelity, rather than limitations of eye-tracking technology, consistent with broader warnings about conflating realism with cognitive validity (Norman et al., 2012; Hamstra et al., 2014).

Underspecified cognitive tasks limited interpretability. Studies that lacked explicit task decomposition or failed to identify which information sources were relevant at specific points in time could not meaningfully map gaze behavior to cognitive processes. In these cases, gaze metrics were often reported descriptively without specifying the cognitive function being inferred, undermining construct validity and interpretability (Vicente, 1999; Holmqvist et al., 2011).

Cognitively shallow XR environments produced engagement-like gaze patterns with limited diagnostic value. When uncertainty was minimized or decision-making demands were reduced, gaze behavior often reflected exploratory viewing rather than task-structured attentional allocation. Such patterns are consistent with eye-movement research showing that gaze behavior differs qualitatively depending on whether viewing is goal-directed versus exploratory (Rayner, 1998; Holmqvist et al., 2011).

A lack of system context undermined inference. Studies that did not log system state, automation behavior, or task events alongside gaze data could not distinguish attentional failures attributable to the human operator from those induced by interface design or automation coupling. Because cognition is distributed across people and artifacts in operational systems, interpretability requires linking gaze to representations and system dynamics

(Hollnagel & Woods, 2005; Parasuraman & Rizzo, 2007). In these cases, eye tracking did not merely fail to add value; it risked misleading interpretation, particularly when gaze was used to evaluate trainees rather than to evaluate whether the simulation preserved the intended cognitive demands (Hamstra et al., 2014; Qian et al., 2025).

FINDINGS AND IMPLICATIONS

This meta-analysis yielded four empirically grounded findings that clarify when and why eye tracking produces interpretable results in simulation-based training environments. Together, these findings indicate that cognitive fidelity, rather than simulator realism or sensor quality, is the primary determinant of gaze– performance relationships.

Across the studies analyzed, simulations with high cognitive fidelity produced moderate-to-large and stable effects (mean $r = .48$), whereas low-fidelity simulations produced weak and highly heterogeneous effects (mean $r = .07$), independent of technological sophistication. This structured heterogeneity indicates that inconsistencies in the literature arise primarily from variation in task design rather than from limitations of eye tracking itself (Hamstra et al., 2014; Vicente, 1999). Effect size variability decreased systematically as cognitive fidelity increased, reinforcing the conclusion that heterogeneity was not random but reflected whether simulations preserved the information-processing demands of real work (Qian et al., 2025).

Eye tracking yielded interpretable results only when gaze measures were anchored to **task-relevant information structures**. Studies relying on global gaze metrics in cognitively shallow simulations consistently failed to differentiate expertise or predict performance, consistent with foundational eye-movement research showing that gaze behavior is meaningful only under goal-directed task constraints (Rayner, 1998; Holmqvist et al., 2011). When simulations imposed meaningful cognitive demands (requiring learners to allocate attention dynamically, anticipate future system states, and manage uncertainty) eye tracking captured stable, task-relevant behavioral signatures associated with expert performance, such as anticipatory monitoring and selective attention to diagnostically relevant cues (Endsley, 2000; Wickens, 2008).

Conversely, when simulations lacked cognitive fidelity, gaze behavior primarily reflected exploratory viewing, engagement, or idiosyncratic scanning patterns rather than goal-directed attentional control. From a neuroergonomic standpoint, this does not represent a measurement failure but an ecologically valid signal that the task environment does not sufficiently constrain cognition to reveal stable behavioral regularities. In such cases, eye tracking accurately reveals the absence of cognitively meaningful structure rather than a deficiency in the learner.

A further finding concerns the importance of system context. Eye tracking frequently hindered inference when task events, automation states, or representational changes were not logged alongside gaze data. Without this context, gaze patterns risked being misattributed to individual attentional failures rather than to design-induced constraints or automation behavior, particularly in joint cognitive systems where cognition is distributed across people and artifacts (Hollnagel & Woods, 2005; Parasuraman & Rizzo,

2007). This supports a core neuroergonomic principle: *cognitive processes in real-world work are expressed through behavior constrained by system design*, not through isolated internal states (Parasuraman & Rizzo, 2007; Vicente, 1999). Eye movements do not index cognition directly; they reflect how attention, perception, and decision making are shaped by task representations, information architecture, and environmental constraints embedded within joint cognitive systems.

These results lead to two clear design implications for simulation-based training research and development: 1) cognitive work must be deliberately engineered before it can be meaningfully measured, and 2) eye tracking should follow design stabilization rather than early prototype exploration. Simulations must preserve distributed information sources, temporal dependencies, and uncertainty that require anticipatory monitoring. Introducing eye tracking before these demands are specified often yields weak or misleading results, not because the measure is ineffective, but because the simulation does not yet elicit cognitively relevant behavior. Cognitive engineering methods such as cognitive work analysis, cognitive task analysis, and expert elicitation are therefore prerequisites for valid neuroergonomic measurement (Vicente, 1999; Hollnagel & Woods, 2005). When task structure, automation behavior, or interface representations are still in flux, gaze behavior reflects transient design artifacts rather than stable cognitive strategies, increasing heterogeneity and undermining inference. The present moderator analysis shows that effect sizes strengthen and stabilize only when simulations reach sufficient cognitive maturity, indicating that eye tracking is best suited for comparative evaluation of design alternatives—for example, assessing whether changes in information layout, automation transparency, or scenario structure elicit attentional patterns more consistent with expert performance. Used in this way, eye tracking supports evidence-based refinement rather than exploratory speculation.

These implications position eye tracking as a late-stage validation and refinement tool within the simulation development lifecycle. Applied prematurely, it risks reinforcing design errors; applied appropriately, it provides a sensitive neuroergonomic diagnostic of cognitive fidelity.

CONCLUSION

This meta-analysis demonstrates that eye tracking yields interpretable, reproducible insights in simulation-based training only when environments preserve the cognitive structure of real work. Cognitive fidelity emerged as the dominant moderator of gaze–performance relationships, accounting for both effect magnitude and heterogeneity across studies. High cognitive fidelity simulations consistently produced stronger, more stable effects, whereas low fidelity simulations produced weak and highly variable results regardless of technological sophistication.

These findings reposition eye tracking as a neuroergonomic diagnostic of cognitive fidelity, rather than as a general-purpose measure of attention, engagement, or learning. Gaze behavior reflects how cognitive processes unfold through interaction with system representations, making it particularly well suited for evaluating whether training environments elicit the intended

forms of attention, monitoring, and decision making. Importantly, this work also clarifies the conditions under which eye tracking can hinder inference, highlighting the need for deliberate sequencing in simulation design and evaluation. Cognitive work must be engineered before it can be measured, and measurement must follow design stabilization rather than precede it.

By grounding eye tracking in cognitive engineering principles and empirical moderator evidence, this study opens several avenues for future applied research. These include extending the present synthesis to team-level visual coordination, integrating eye tracking with other neuroergonomic measures under high cognitive fidelity conditions, and using gaze behavior as a comparative metric for evaluating AI-generated and adaptive training environments. Together, these directions position eye tracking not as an endpoint, but as a critical component of a broader research agenda aimed at designing, validating, and scaling cognitively faithful training systems for complex, real-world work.

REFERENCES

- Alessi, S. M. (2000). Simulation design for training and assessment. *Journal of Educational Computing Research*, 22(3), 249–276. <https://doi.org/10.2190/9EW7-04X1-0EHB-YK0M>.
- Alexander, A. L., Brunye, T., Sidman, J., & Weil, S. A. (2005). From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in PC-based simulations and games. *DARWARS Training Impact Group Report*.
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), 32–64. <https://doi.org/10.1518/001872095779049543>.
- Endsley, M. R. (2000). Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 3–32). Lawrence Erlbaum Associates.
- Ericsson, A. K., & Pool, R. (2016). *Peak: Secrets from the new science of expertise*. Houghton Mifflin Harcourt.
- Hamstra, S. J., Brydges, R., Hatala, R., Zendejas, B., & Cook, D. A. (2014). Reconsidering fidelity in simulation-based training. *Academic Medicine*, 89(3), 387–392. <https://doi.org/10.1097/ACM.0000000000000130>.
- Hays, R. T., & Singer, M. J. (1989). *Simulation fidelity in training system design: Bridging the gap between reality and training*. Springer.
- Hochmitz, I., & Yuviler-Gavish, N. (2011). Physical fidelity versus cognitive fidelity training in procedural skills acquisition. *Human Factors*, 53(5), 489–501. <https://doi.org/10.1177/0018720811412777>.
- Hollnagel, E., & Woods, D. D. (2005). *Joint cognitive systems: Foundations of cognitive systems engineering*. CRC Press.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
- Klein, G., Ross, K. G., Moon, B., Klein, D. E., Hoffman, R. R., & Hollnagel, E. (2010). Macrocognition. *IEEE Intelligent Systems*, 18(3), 81–85. <https://doi.org/10.1109/MIS.2003.1200735>.

- Norman, G., Dore, K., & Grierson, L. (2012). The minimal relationship between simulation fidelity and transfer of learning. *Medical Education*, 46(7), 636–647. <https://doi.org/10.1111/j.1365-2923.2012.04243.x>.
- Parasuraman, R., & Rizzo, M. (Eds.). (2007). *Neuroergonomics: The brain at work*. Oxford University Press.
- Qian, X., et al. (2025). Why do we need high-fidelity synthetic eye movement data and how should they look? [Preprint]
- Rahwan, I., et al. (2019). Machine behaviour. *Nature*, 568, 477–486. <https://doi.org/10.1038/s41586-019-1138-y>.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422. <https://doi.org/10.1037/0033-2909.124.3.372>.
- Salas, E., Tannenbaum, S. I., Kraiger, K., & Smith-Jentsch, K. A. (2012). The science of training and development in organizations. *Psychological Science in the Public Interest*, 13(2), 74–101. <https://doi.org/10.1177/1529100612436661>.
- Salas, E., Wilson, K. A., Burke, C. S., & Priest, H. A. (2009). Using simulation-based training to improve patient safety. *Joint Commission Journal on Quality and Patient Safety*, 35(9), 481–490.
- Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2021). Organizational decision-making structures in the age of artificial intelligence. *California Management Review*, 63(3), 66–83. <https://doi.org/10.1177/0008125621995145>.
- van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning. *Educational Psychology Review*, 17(2), 147–177. <https://doi.org/10.1007/s10648-005-3951-0>.
- Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*. CRC Press.
- Wickens, C. D. (2008). Situation awareness: Review of Mica Endsley's 1995 articles. *Human Factors*, 50(3), 397–403. <https://doi.org/10.1518/001872008X288420>.