Maritime Policy & Management
The flagship journal of international shipping and port research

ISSN: 0308-8839 (Print) 1464-5254 (Online) Journal homepage: www.tandfonline.com/journals/tmpm20

Studying the impact of contextual factors on language barriers in maritime safety: a structural equation modeling approach

Peng Liu, Rafi Ullah Khan, Muhammad Afzaal, Mujtaba Asad, Xiangtao Du & Zaili Yang

To cite this article: Peng Liu, Rafi Ullah Khan, Muhammad Afzaal, Mujtaba Asad, Xiangtao Du & Zaili Yang (04 May 2026): Studying the impact of contextual factors on language barriers in maritime safety: a structural equation modeling approach, Maritime Policy & Management, DOI: 10.1080/03088839.2026.2660896

To link to this article: https://doi.org/10.1080/03088839.2026.2660896

Published online: 04 May 2026.

MARITIME POLICY & MANAGEMENT
https://doi.org/10.1080/03088839.2026.2660896

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Studying the impact of contextual factors on language barriers in maritime safety: a structural equation modeling approach

Peng Liu[a], Rafi Ullah Khan[b], Muhammad Afzaal[c], Mujtaba Asad[d], Xiangtao Du[e] and Zaili Yang[b]

[a]School of Foreign Languages, Hainan Tropical Ocean University, Sanya, China; [b]Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool, UK; [c]Institute of Corpus Studies and Applied Linguistics, Shanghai International Studies University, Shanghai, China; [d]Institute of Image Processing and Pattern Recognition, Department of Automation, Shanghai Jiao Tong University, Shanghai, China; [e]College of Foreign Languages, Shanghai Maritime University, Shanghai, China

ABSTRACT

Communication failures, particularly language barriers, are critical contributors to maritime accidents, yet their impact remains under-explored. This study quantitatively assesses the effect of language and communication barriers on maritime accidents, investigating the interactions between human, operational, and organizational factors. Data from 515 maritime accident reports (2012–2022) was processed using Natural Language Processing (NLP) techniques, extracting observed variables including language proficiency, accent issues, cognitive load, and navigational complexity. Factor analysis and Structural Equation Modeling (SEM) were used to identify latent variables influencing accident risk. SEM was employed to analyze the relationships between latent variables, human factors, operational context, and organizational conditions, and the observed. Results show that human factors, especially stress and fatigue, significantly impact accident causation by compromising communication and navigation safety. Language proficiency (Low) ($\lambda = 0.91$, SE = 0.03, z = 30.33) was revealed as the most critical factor; however, operational aspects like navigational complexity ($\lambda = 0.93$, SE = 0.04, z = 23.25), time pressure, and organizational constraints exacerbate communication breakdowns and increase accident likelihood. The SEM (CB-SEM (Covariance-Based-SEM), WLSMV (Weighted Least Squares Mean and Variance Adjusted)) demonstrated strong explanatory power (R2 = 0.845) and excellent fit (RMSEA = 0.024, SRMR = 0.012, CFI = 0.962, TLI = 0.971). The study highlights the need for improved training, communication protocols, and language proficiency to reduce maritime accident risks.

ARTICLE HISTORY
Received 21 May 2025
Accepted 6 April 2026

KEYWORDS
Language barriers; maritime safety; Structural Equation Modelling (SEM); human factors; operational context

CONTACT Zaili Yang Z.Yang@ljmu.ac.uk; Raff Ullah Khan r.u.khan@ljmu.ac.uk Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool L3 3AF, UK.

1. Introduction

Maritime transportation is a high-risk industry that relies on effective human performance and communication to ensure safety (Crestelo Moreno et al. 2022; Khan et al. 2022). A vast majority of maritime accidents are attributed to human factors; studies estimate that between 75% and 96% of incidents are caused by human error (Albert 2024; Zhou et al. 2025). These human factors encompass a wide range of issues, including cognitive overload, fatigue, poor communication, and flawed decision-making (Antão and Guedes Soares 2019; Chauvin 2011; Khan et al. 2021; H. Wang et al. 2025). In particular, communication failures on the bridge and between vessels have emerged as a critical safety concern in the global shipping industry (Boström 2020a, 2020b; Ahmmed 2020). Modern merchant ships are typically manned by multicultural, multilingual crews, as approximately 80% of seafarers do not speak English as their first language (2020). While English is the lingua franca of the seas and the International Maritime Organization's (IMO) Standard Marine Communication Phrases (SMCP) provides a standardized vocabulary, language barriers and miscommunication persist (Boström 2021; Haryani, Januarius, and Issy 2023; John, Brooks, and Schriever 2017; Noble 2015). Communication breakdowns have been identified as contributing factors in *up to one-third of ship accidents* according to industry analyses (2020), underscoring the need for focused research on this issue.

Despite advances in technology and stricter regulations, incidents continue to occur where the root cause can be traced to a failure in communication. High-profile maritime accidents illustrate the deadly consequences of language barriers. For example, investigators of the *Cosco Busan* collision (San Francisco Bay, 2007) found that a major cause was the lack of effective communication between the American pilot and the Chinese captain; many crew members had limited English proficiency and could not fully understand critical instructions (2020; Bošnjak et al. 2021). In another tragic case, the *Scandinavian Star* ferry fire (1990) resulted in 161 fatalities, and post-incident analysis highlighted that the multinational crew's inability to communicate in a common language severely hampered emergency response and evacuation efforts (2020; Koenig and Schultz 2016). These cases, among numerous others, demonstrate that language and communication problems are not merely inconveniences; they are direct threats to maritime safety. The IMO has long recognized this, mandating English for bridge communications and promoting SMCP to reduce ambiguity (L. Liu et al. 2026; Yin et al. 2025). However, the persistence of communication-related casualties suggests that deeper, contextual factors influence when and how language barriers lead to accidents.

Given the multicultural nature of today's seafaring workforce and the complex, high-pressure environments in which they operate (Yu et al. 2025), it is crucial to understand how contextual factors exacerbate or mitigate language barriers on the front lines of maritime safety. The present study examines how such factors, including cognitive load, fatigue, stress, navigational complexity, and time pressure, interact with communication challenges to contribute to accident causation. While existing research has provided valuable insights into human error and cross-cultural communication, these aspects have often been treated in isolation (Yu et al. 2025). Consequently, there remains a limited understanding of how multiple stressors and human factors collectively amplify the risks associated with miscommunication at sea.

To address these limitations, this study focuses on two interrelated gaps. First, existing research has not yet provided an integrated quantitative model that jointly represents linguistic barriers, human-state factors (fatigue, stress, cognitive load), operational context (navigational complexity, weather, time pressure), and organizational conditions (training, safety culture, crew composition) within a single causal framework. Prior work typically treats these dimensions separately or qualitatively, which makes it difficult to quantify their combined and mediated effects on accident risk. Second, although SEM has been used in maritime safety studies, it has not been applied to models that include explicit communication and language constructs, nor has it been systematically linked to variables extracted from large corpora of investigation reports using NLP and LLM-based tools. As a result, there is limited empirical evidence on how narrative-derived indicators of language/communication failure, human states, and context interact structurally to influence accident outcomes.

The present study addresses these gaps by combining NLP-driven variable extraction from 515 accident investigation reports with a covariance-based SEM (WLSMV) that integrates linguistic, human, operational, and organizational factors into a single multi-latent model of accident risk. This design allows us to (i) derive indicators and constructs directly from real accident narratives rather than from small-scale surveys alone, (ii) estimate the relative contributions and pathways linking language barriers and contextual human factors to accident severity, and (iii) quantify how organizational conditions shape these pathways. By doing so, the study provides a distinct contribution to maritime safety research (Hu et al. 2019; Othman and Adolf Naintin 2016; Vivek et al. 2024; Wahid et al. 2024; Xu, Xiao, and Jiang 2021): it connects unstructured narrative evidence to a rigorously specified SEM, generating both explanatory insight and policy-relevant levers for reducing communication-related accident risk.

The following literature review (Section 2) elaborates on the existing state of knowledge regarding maritime communication failures in multilingual settings, the role of human and operational factors, and the methodological opportunities that this study seeks to build upon. The literature review is followed by the methodology (Section 3), which encompasses the adopted approach, and then results and discussions (Section 4) present the findings and evaluate the potential dynamics and impacts. Next is the practical implications (Section 5), which showcases the potential insinuations of the findings, and Conclusions (Section 6) conclude the study.

2. Literature review

2.1. Communication challenges in multilingual maritime operations

Shipping has become a truly global enterprise, and vessels are often crewed by seafarers of diverse nationalities and cultural backgrounds. This diversity brings benefits but also significant communication challenges. Research on maritime human factors has consistently found that the lack of a common language on board can turn a ship into a high-risk work environment (Boström 2020a, 2020b; 2021). Crew members who do not share a native tongue may interpret words, gestures, and even safety protocols differently, increasing the potential for misunderstandings (Ahmmed 2020; Argüelles et al. 2021; Bialyk 2017a, 2017b; Ana 2011). Even when English is used as the common language, varying levels of proficiency and heavy accents can impede clear dialogue. Communication is itself a complex psycho-social process involving not just words but also tone, context, and body language. Barriers to effective communication can stem from physical noise (e.g. engine noise), technological issues (e.g. poor radio signal), or physiological and psychological factors (stress, anxiety) (L. Liu et al. 2026; Yin et al. 2024, 2025). Notably, in high-stress situations, people often revert to their mother tongue unconsciously. This means that during an emergency or time-critical maneuver, a crew member might issue a warning or request in their native language, potentially leaving others on the bridge confused at exactly the worst time. Such scenarios have been reported in accident investigations; for instance, bridge audio transcripts occasionally show officers switching languages when alarms are sounding or when under extreme pressure (2020).

To mitigate language difficulties, the maritime industry has put in place standards and training for *Maritime English*. The IMO formally adopted Seaspeak and later the Standard Marine Communication Phrases as tools to enforce clear, concise English terminology for common maneuvers and shipboard commands (Y. Li 2019; Pejaković 2014). These measures have undoubtedly improved communications to some extent. However, adherence to SMCP is not universal, and real-world communications often deviate from the textbook phrases, especially in novel or emergency scenarios. Multilingual crews still face subtle challenges: differences in culture can affect communication styles (for example, some cultures may be less inclined to question a superior's decision or may use indirect speech, which can be misinterpreted). A study observes that integrating a mixed crew remains a *maritime safety challenge*, noting that ship operators historically have been wary of crew heterogeneity and sometimes prefer more culturally homogeneous teams (Horck 2008). Nevertheless, modern manning practices and labor markets make diverse crews the norm rather than the exception. Thus, the industry must confront the safety issues arising from multicultural and multilingual settings. Recent studies continue to document incidents where language or cultural differences played a role in errors. For example, a case study applied automatic text mining to 344 accident investigation reports and identified numerous 'unsafe acts' linked to communication issues, demonstrating that miscommunication remains a prevalent causal factor in accidents across different types of vessels (Shi et al. 2022). In sum, while English proficiency and standardized phrases form the foundation of maritime communication, the realities of a multilingual crew environment introduce persistent risks that require ongoing attention in both research and practice. This sets the stage for examining how communication failures translate into accidents, and under what conditions these failures are most likely to occur.

2.2. Communication failures and contextual human factors in maritime safety

Poor communication, within a bridge team, between ship and shore (e.g. VTS or pilot), or between vessels, has been repeatedly implicated in maritime accidents. In accident report analyses, communication problems are often categorized as a type of human or 'personnel' factor that serves as a precondition for unsafe acts (Satler, Tanja, and Jurkovič 2024). For instance, a seminal analysis of collision cases highlighted that deficits in inter-ship communication and breakdowns in Bridge Resource Management were among the leading preconditions to collisions at sea (Chauvin et al. 2013). This study, which adapted the Human Factors Analysis and Classification System (HFACS) to maritime collisions, found that while decision errors by individuals were the top immediate cause of accidents, those decisions were frequently shaped by upstream communication deficiencies, such as unclear passing arrangements between vessels or misunderstandings between pilots and bridge teams (Chauvin et al. 2013). In restricted waters, especially, Chauvin et al. noted a strong correlation between accidents and inter-ship communication problems, often involving vessels with different language backgrounds or where standard protocols (like repeating back instructions) were not followed (Chauvin et al. 2013). This underscores that even when all crew speak a common language, communication across vessels (over VHF radio, for example) can suffer from imprecise language or non-standard phraseology, leading to confusion about intentions and actions. A Safety Board investigation into a collision between a bulk carrier and a tanker similarly concluded that the VHF radio exchanges were 'not precise or simple,' and suggested that had the officers used standard phrasing, the risk of confusion would have been greatly reduced (2020).

Multiple studies and incident statistics attest to the outsized role of communication failures in maritime casualties. According to one review, communication deficiencies (such as misunderstandings, lack of timely information exchange, or ambiguous orders) contribute to roughly 40% of accidents at sea (Ma et al. 2022; Yin et al. 2024, 2025). In some domains of operation, the figures are even more stark; an analysis of offshore industry accidents found that nearly 73% of incidents had some form of communication failure as a contributing factor, particularly failures to clarify or confirm information (Ziegler 2024). Common communication breakdowns include: failure to use closed-loop communication (i.e. not repeating back and verifying instructions), language proficiency issues (crew not fully understanding an order but failing to ask for clarification), and contextual misunderstandings (such as misinterpreting a colloquial phrase or a non-standard command). A case study of maritime accidents found that poor communication by seafarers was a main factor leading to a catastrophic accident (a boiler explosion on a tanker), demonstrating how a simple miscommunication can cascade into a disaster (Celik and Cebi 2009). Onboard, ineffective communication erodes teamwork and trust. Another study noted that open and direct communication is one of the 'pillars of a healthy organization' on a ship, and that inefficient or inadequate communication invariably creates unsafe conditions (Hasanspahić et al. 2021). Examples from incident reports include crew members not sharing critical information: berthing plans not discussed with the team, watch officers failing to inform the master of urgent navigational needs, or engineers not alerting the bridge about technical problems (Hasanspahić et al. 2021).

It is also important to recognize the organizational and cultural context around communication. Shipping companies via the ISM Code are required to establish clear communication channels and protocols, but a company's safety culture determines how comfortable junior crew feel in speaking up. A vessel with a rigid, top-down culture might discourage questioning or repeating orders for confirmation, which can silence vital communication. On the other hand, a crew that practices assertive closed-loop communication, where every instruction is acknowledged and verified, can catch and correct miscommunications before they cause harm. The UK

Maritime Coastguard Agency includes communication as one of the 'Deadly Dozen' human factors in maritime safety, emphasizing training in effective information exchange and assertiveness to reduce errors (Simões-Marques, Frias, and Água 2021).

Maritime operations frequently take place in challenging conditions that can tax the crew's cognitive and physical capacities. These contextual factors, high workload, fatigue, psychological stress, complex navigation situations, and time urgency, can dramatically worsen the effect of any language barrier or communication hiccup. Recent research has reinforced how excessive Cognitive Load on mariners contributes to errors and accidents. Cognitive load refers to the mental effort required to monitor and respond to a situation. On a busy ship's bridge, officers must continuously process information from radar/ARPA, ECDIS displays, engine controls, radio communications, and visual lookout. When confronted with heavy traffic, poor visibility, or an alarm situation, the officer's mental workload can spike to a point of overload. Studies have linked such high cognitive demand to breakdowns in performance: for example, an analysis found that about 75% of marine incidents in one merchant fleet were attributed to human erroneous actions mainly induced by high cognitive load on the officer of the watch (Miklody et al. 2016; Žagar et al. 2020). Under these conditions, attention and situational awareness are diminished. Importantly, communication tends to suffer as well; an overtaxed operator might miss a radio call, fail to convey a message, or misunderstand a critical piece of information because their attention is divided. Empirical experiments in ship simulators support this: as mental workload increases (due to dense traffic, bad weather, or additional tasks), officers' reaction times slow, and the incidence of miscommunications and rule violations rises. In essence, high cognitive load can directly lead to misunderstandings and errors, which is especially dangerous if those misunderstandings involve navigation instructions or collision-avoidance communications (Maglić et al. 2020). Likewise, diminished situational awareness often accompanies a heavy workload; when an officer is overloaded, they may not fully notice or absorb communications from team members, leading to crucial information gaps (Albert 2024).

Fatigue is another pervasive issue that significantly amplifies communication problems. Seafarers work in a demanding environment with long hours, night shifts, and often insufficient rest, all of which contribute to chronic fatigue. The link between fatigue and accidents is well-documented. Chronic sleep deprivation and tiredness impair cognitive function, reduce alertness, and degrade both decision-making and communication skills (Albert 2024; Setiawan and Razak Abdul Hadi 2025). An exhausted crew member is more prone to speak indistinctly, forget to relay information, or fail to comprehend an instruction. Recent studies confirm that fatigue remains widespread despite regulatory limits on work hours. As a result, fatigue was identified as a contributing factor in nearly 20% of marine incidents (Vagias 2010). Fatigue's effect on communication can be both direct and indirect: directly, a fatigued person might simply miss words or phrases in a conversation, or lack the energy to double-check that their message was understood. Indirectly, fatigue lowers overall vigilance; one may not realize a misunderstanding has occurred until it's too late. Furthermore, fatigue often co-occurs with other risk factors; for example, a fatigued officer under time pressure is an especially dangerous combination, as their ability to communicate clearly and make sound judgments is severely compromised (Vagias 2010).

Closely related to workload and fatigue is stress, which can stem from emergencies, time pressure, or even interpersonal tensions on board. High stress levels trigger the body's 'fight or flight' response, which can narrow an individual's attention and impede complex cognitive processing. In maritime emergencies, stress is unavoidable, but it can lead to detrimental communication outcomes. Under acute stress, individuals may revert to their native language or default communication habits (as noted earlier) (2020), and they may omit important details or fail to listen effectively. Research in maritime human factors has observed that stressors like time pressure, bad weather, and high-stakes decision-making often impair human performance, including the ability to communicate and coordinate (Arenius, Athanassiou, and Sträter 2010; M. H. Lützhöft and Dekker 2002; M. Lützhöft, Grech, and Porathe 2011; Ronca et al. 2023). For example, a bridge team under extreme time pressure to avoid a collision may experience tunnel vision, where each person is fixated on their immediate task and communication becomes curt or incomplete. The presence of stress can significantly degrade cognitive functioning and attentional focus, increasing the risk of accidents (M. H. Lützhöft and Dekker 2002; M. Lützhöft, Grech, and Porathe 2011). Time Pressure is particularly hazardous: compressed decision windows encourage shortcutting standard protocols and verification, amplifying the consequences of language-related ambiguity (Committee 2023). A tragic example was the Sanchi tanker collision (2018), where time-critical decisions had to be made in a congested area; investigators noted that the hurried exchange of information contributed to confusion about each vessel's next move. While technology (like AIS and ARPA) provides some mitigation by conveying intent, nothing can fully replace clear, unambiguous crew-to-crew communication in such moments. Therefore, navigational complexity and time urgency are important contextual variables to consider; they represent moments when communication barriers are most likely to lead to an accident if not properly managed.

The literature converges on the conclusion that communication failures, magnified by language barriers, are not isolated faults but the behavioral expression of interacting contextual conditions (Navigational Complexity, Weather Conditions, Time Pressure) and human states (Cognitive Load, Stress and Fatigue), modulated by organizational factors (Crew Training, Safety Culture). This is consistent with accident statistics (2020), causal analyses (Chauvin et al. 2013), and human-factors studies (Albert 2024; Arenius, Athanassiou, and Sträter 2010; M. H. Lützhöft and Dekker 2002; M. Lützhöft, Grech, and Porathe 2011; Maglić et al. 2020; Miklody et al. 2016; Ronca et al. 2023; Setiawan and Razak Abdul Hadi 2025; Vagias 2010; Žagar et al. 2020).

2.3. Analytical approaches to studying maritime communication and safety

Researchers have employed various methods to study language barriers and communication issues in maritime safety, ranging from qualitative case studies to advanced statistical modeling. Traditional incident analyses often use human factors frameworks like HFACS to classify the underlying causes in accidents (Khan et al. 2021). HFACS-based studies consistently underscore communication problems as a prominent 'precondition for unsafe acts' in maritime casualties (Hasanspahić et al. 2021; Satler, Tanja, and Jurkovič 2024). For example, Chauvin et al.'s work mentioned earlier used a modified HFACS to categorize errors in collision reports, while Celik & Cebi applied

HFACS to a tanker explosion investigation and identified poor crew communication as a key factor precipitating the accident (Celik and Cebi 2009). While such frameworks qualitatively identify where communication fits into the causal chain, they do not quantify the strength of its impact or the interrelations with other factors. This is where statistical approaches like factor analysis and Structural Equation Modeling (SEM) come into play. SEM is a powerful technique for testing hypotheses about relationships among latent constructs (unobserved variables) by using measurable indicators (Byrne 2013; Hu et al. 2019). In the context of maritime safety research, SEM allows investigators to model complex cause-effect networks—for instance, how crew language proficiency (latent variable) and workload (latent variable) together influence the occurrence of communication errors, which in turn affect accident likelihood. Early applications of SEM in safety science demonstrated its utility; using SEM to reveal the mechanisms through which contributory factors of unsafe work behavior influence safety outcomes (D.-C. Seo 2005; H.-C. Seo et al. 2015). More recently (Hu et al. 2019), developed a structural equation model to analyze causal factors in marine traffic accidents, integrating HFACS categorizations with SEM to illuminate the dependencies among human errors, unsafe acts, and accident occurrence. Their approach confirmed that SEM could quantify the influence of factors like decision errors, violations, and communication lapses on accident outcomes, providing a more nuanced understanding beyond simple correlation. The advantage of SEM lies in its ability to handle multiple variables and pathways simultaneously, a fitting approach for our study, which posits that contextual human factors indirectly affect accident risk *through* their impact on communication effectiveness. We will follow this path by using factor analysis to reduce a broad set of human and operational variables into meaningful dimensions, and then apply SEM to test how those dimensions relate to communication breakdowns and safety outcomes.

Another noteworthy development in the literature is the use of Natural Language Processing (NLP) and text mining on maritime accident reports. Traditional analyses of accident reports are labor-intensive, requiring experts to read and code narratives. Recent studies leverage NLP tools to extract common themes and factors from large collections of investigation reports. For example (Shi et al. 2022), employed an automated text mining approach on hundreds of investigation reports from international databases to identify frequent contributory factors and sequences of unsafe acts. This innovative approach can uncover patterns not immediately obvious to human analysts, such as recurring language that indicates a communication problem (e.g. 'language difficulty,' 'misheard instructions,' 'no response received'). Similarly, other researchers have mined textual data to find causal relations: one data-driven study used text mining to reveal that *'lack of communication'* and *'fatigued ship handling'* often co-occur in collision narratives (Y. Wang and Fu 2022), suggesting an interplay between these factors. The combination of NLP with human factors analysis opens a promising avenue—by processing thousands of lines of text from reports, one can systematically quantify how often certain factors (like language issues, fatigue, time pressure) are mentioned together or linked causally. Such findings strongly inform the modeling stage (for instance, guiding which factors to include in an SEM). Large Language Models (LLMs) represent the cutting edge of NLP, capable of interpreting complex contexts and even inferring implicit information from text. While still emerging in academic research, LLMs can be used to parse accident reports in a more nuanced way, for example, recognizing when a report implicitly attributes confusion to language problems even if not stated explicitly. In this study, we harness NLP techniques (supported by LLM-based parsing) to extract variables and indicators from narrative reports, creating a rich dataset for factor analysis. This approach is aligned with the direction of contemporary research, which increasingly uses data-driven analysis of investigation reports to supplement expert judgment (Shi et al. 2022). By doing so, we ensure that our factor identification is grounded in real-world evidence of what goes wrong in accidents, rather than just theoretical considerations.

2.4. Theoretical linkages and conceptual framework

Sections 2.1–2.2 establish that language/communication issues, as well as human factors, co-occur in incidents. Here, we make the linkage explicit and organize the variables into a coherent mechanism that culminates in Accident Risk.

2.4.1. Operational context → human states

Navigational Complexity, Weather Conditions, and Time Pressure increase perceptual and decision demands on the bridge team. This raises Cognitive Load (more cues to track, faster integration required) and elevates Stress and Fatigue (compressed time windows, degraded vigilance). These human states are the proximal conditions under which communication becomes error-prone.

2.4.2. Language barriers → processing demand and violation risk

Low Language Proficiency and more severe Accent/Pronunciation increase the decoding effort required to understand messages, thereby increasing Cognitive Load and directly raising the chance of a Communication Role Violation (skipping role/rule-based protocols, incomplete/unclear statements, missed confirmations). Multilingual Crew Composition further increases coordination and code-switching demands, which elevates Cognitive Load and heightens the sensitivity of communication to proficiency/accent differences.

2.4.3. Human states → violation mechanism

Higher Cognitive Load and higher Stress and Fatigue degrade message production and comprehension (slips, omissions, delayed or ambiguous responses), increasing the likelihood of a Communication Role Violation.

2.4.4. Experience as a protective human factor

Greater Experience with multinational crews and standardized protocols reduces Cognitive Load under difficult contexts (better cue prioritization and schema use) and directly lowers Communication Role Violation by improving when/what/how to communicate.

2.4.5. Organizational enablers as buffers

Stronger Crew Training (on standardized communication practices) and a stronger Safety Culture (emphasizing closed-loop communication and consistent language use) both reduce Communication Role Violation directly and buffer the translation of high Cognitive Load, Stress, and Fatigue, and language barriers into violations (by enforcing verification and shared terminology).

2.4.6. Proximal mechanism to outcome

Communication Role Violation is the immediate behavioral mechanism through which context, human states, and language barriers influence safety; higher violation likelihood increases Accident Risk.

2.5. Scope and significance

Based on the existing literature, it is clear that communication failures in multilingual bridge teams are a major safety concern, shaped by a combination of human and contextual factors rather than language alone. Studies have documented how language barriers arise in everyday shipboard communication and how they feature in accident narratives (2020; Chauvin et al. 2013; Yin et al. 2024). Other work has shown that fatigue, stress, and high workload frequently appear in the background of incidents, eroding attention, situation awareness, and decision quality (Albert 2024; Miklody et al. 2016; Žagar et al. 2020). Investigations of major casualties further demonstrate that breakdowns in communication rarely occur in isolation; they tend to coincide with situational stressors, cognitive overload, and organizational shortcomings in training, supervision, or safety culture (Hetherington, Flin, and Mearns 2006).

When this body of work is viewed together, two specific gaps emerge. First, accident studies and safety models seldom bring language barriers, human state factors, operational conditions, and organizational features into a single, quantitative causal structure. Communication issues are often described qualitatively or discussed alongside cultural differences, but prior research has not systematically modeled how linguistic barriers interact with fatigue, time pressure, or navigational complexity to produce accidents (Boström 2021; Horck 2008; Mallam, Nazir, and Sharma 2020). In practice, this means that we know 'language matters' and 'fatigue matters,' but we lack an integrated picture of how these factors combine along a causal chain from bridge conditions to accident risk. Second, while structural equation modeling has been used to study maritime safety topics such as unsafe acts (Hu et al. 2019), pilots' risky behaviour (Xu, Xiao, and Jiang 2021), and seafarers' mental health (Baygi et al. 2022), these SEM applications have not embedded linguistic and communication variables as explicit constructs, nor have they been calibrated against large sets of real accident reports. At the same time, a newer line of research uses NLP to extract causes and contributory factors from narrative safety reports (Lan, Wang, and Zhang 2024; D. Liu and Cheng 2024; Ricketts et al. 2023), but these NLP-based indicators have not been systematically coupled with SEM to represent structural pathways and interdependencies among language, human, operational, and organizational dimensions.

Recent work in *Maritime Policy & Management* further underlines both the need and the opportunity for such an integration. One study assesses the readability of maritime accident reports and shows that narratives vary widely in clarity, arguing for methods that can transform free text into more structured, analysable data (Goerlandt and Liu 2024). Our NLP pipeline responds directly to this challenge by extracting harmonised indicators from accident narratives and preparing them for structural modeling. Another contribution demonstrates the potential of machine-learning models for predicting maritime risk *(Knapp and van de Velden 2024)*. In contrast, our approach is not primarily predictive; instead, it opens up the mechanisms behind risk by linking language barriers, human and organizational factors, and operational context in a theory-driven SEM. In parallel, research on safety climate and team dynamics shows how organizational conditions shape seafarers' safety behaviour (Chen et al. 2023); these insights are reflected in our Organizational Factors latent, which connects safety culture, training, and multilingual crew composition to communication outcomes. Further, spatial analyses of accident consequences highlight how risk patterns depend on local context (G. Li et al. 2025); our model operationalises such contextual stressors through indicators like navigational complexity and time pressure and links them explicitly to communication breakdowns and accident risk. Finally, a corpus-based risk assessment of SMCP/Maritime English demonstrates a measurable link between linguistic features and miscommunication risk (L. Liu et al. 2026); in our framework, these language-related risks are embedded within a broader multi-latent structure that culminates in accident risk. Collectively, these studies set a clear stage that this paper extends by combining narrative-derived indicators with an integrated SEM of language, human, operational, and organizational dimensions.

Against this backdrop, the contribution of the present study is to move from fragmented insights to a single, empirically grounded model that quantifies how language and communication problems interact with human state, operational context, and organizational setting to shape accident risk. By using NLP to structure a large corpus of accident reports and then applying factor analysis and SEM, we are able to estimate the relative strength of these pathways using real-world evidence rather than hypothetical scenarios or small-scale surveys. This allows us to answer questions that previous work has only raised qualitatively: for example, whether time pressure amplifies the effect of low language proficiency, or whether fatigue and weak safety culture jointly increase the likelihood that communication failures will escalate into serious accidents.

The practical significance of this research lies in translating these quantified relationships into targeted interventions. If the model shows that time pressure substantially magnifies the effect of communication barriers, procedures for time-critical phases (e.g. collision avoidance, pilotage, emergency response) can be redesigned to enforce stricter, standardized communication protocols. If combinations such as fatigue plus low English proficiency emerge as especially hazardous, policy measures can include minimum rest, language support, or watch composition rules in such situations. In this sense, the study goes beyond restating that 'communication is important' and instead clarifies how important it is relative to other factors, and under which operational conditions it becomes most critical to reinforce. By explicitly addressing the dual gap, lack of integrated structural models that combine linguistic and contextual stressors, and limited SEM validation using large-scale accident data, this work offers a novel, data-driven framework that directly responds to current limitations in the maritime safety literature. Through the joint use of NLP and SEM on real accident reports, the approach remains both data-informed and holistic, reflective of the complexity of modern, multicultural bridge operations that previous studies have only partially captured.

In summary, the existing literature leaves little doubt that language barriers and communication breakdowns are central to maritime safety, especially in multicultural crewing environments. These failures often act as the final trigger in an accident sequence, yet they sit on top of deeper conditions such as cognitive overload, fatigue, operational strain, and organizational weakness. Modern tools such as NLP and SEM now make it possible to untangle these intertwined influences in a systematic way. Building on this foundation, the present study uses narrative-derived, structured data and a multi-latent SEM to examine how contextual human factors shape language barriers and accident risk, with the aim of informing more precise and effective prevention strategies in future maritime safety governance.

3. Methodology

The study's methodology is organized into four main stages, from initial data gathering to model interpretation. The flowchart in Figure 1 illustrates this process step-by-step, emphasizing the Natural Language Processing (NLP) procedures used to convert unstructured text into structured data. Each stage (1–4) flows into the next with one-directional arrows, showing the sequence: Data Collection → NLP & Data Structuring → Statistical Analysis & SEM → Model Evaluation & Interpretation. Key details of each stage are outlined as follows.

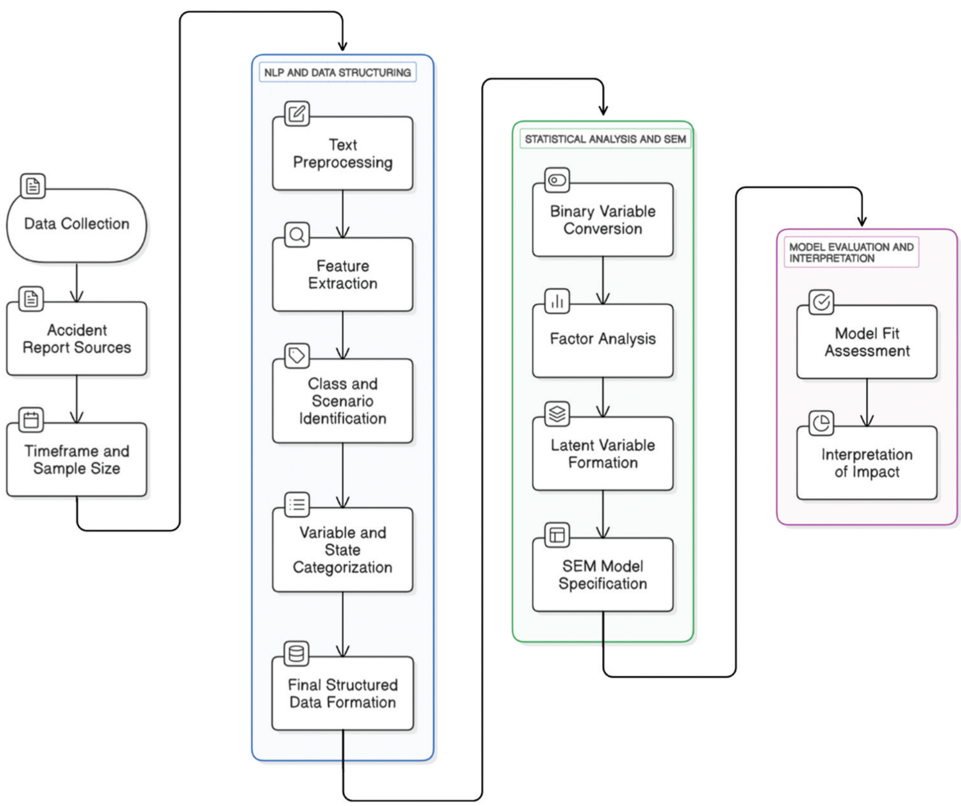

Figure 1. A schematic flowchart of the adopted methodology.

3.1. Data

Human factors, specifically their association with communication failures, are regarded as a major causation factor in maritime accidents. The communication failures, specifically instigated by the lack of language capability, are a prominent yet underexplored domain in maritime safety. The language capability itself is compromised by a range of factors, which in turn are human and environment-centric. Hence, the accidents instigated by language and communication-related issues reveal that, along with these factors, certain other factors either co-instigate these accidents or intensify the role of these factors in accident causation. With the recent organization and technological advancements, there is a plethora of detailed accident investigation reports in the maritime industry. However, specific structured data is scarce. In this domain, structured data entirely centered around language, communication, and relevant causation factors is specifically hard to find.

To this backdrop, the current study accumulates detailed accident investigation reports from diverse sources including Australian Transport Safety Bureau (www.atsb.gov.au), Marine Accident Investigation Branch UK (www.gov.uk/government/organisations/marine-accidentinvestigation-branch), National Transportation Safety Board US (www.ntsb.gov/Pages/home.aspx), Norwegian Safety Investigation Authority (www.nsia.no/Marine/Published-reports), Transportation Safety Board of Canada (www.tsb.gc.ca), China Maritime Safety Administration (https://en.msa.gov.cn/), and the Hong Kong Marine Department published reports (https://www.mardep.gov.hk/en/materials-and-publications/publications/reports/reports-of-marine-department/ereport/index.html). Accident reports between 2012–2022 were gathered and processed to extract the structure data from these reports. NLP techniques were employed to extract language and communication issues-related data from these accident reports. A total of 515 accident reports were finalized to use for the extraction of structured data. Advanced text mining and NLP techniques were utilized through CustomGPTs to transform unstructured maritime accident reports into structured, analyzable datasets by extracting relevant information from qualitative narratives. The steps involved in obtaining the structured data involved defining the key fields, text preprocessing, feature extraction, class and scenario identification, variable and state categorization, and finally structured data formation. Structured variables were produced via an LLM-assisted NLP pipeline configured with a fixed prompt template and extraction schema. The details of the applied approach are provided below.

3.2. Data preprocessing

In the first stage, CustomGPTs were configured to extract structured variables from the free-text investigation reports. The model was guided by a predefined extraction schema that reflected the conceptual domains of interest: language proficiency and communication behavior; human-state factors such as fatigue, stress, and cognitive load; operational conditions such as navigational complexity, and weather; and organizational aspects such as training, safety culture, and crew composition. Defining these target fields in advance ensured that the extraction process was consistent across reports and aligned with the constructs later used in the SEM.

3.2.1. Text preprocessing

Before automated extraction, the accident narratives were cleaned and normalized to support reliable NLP operations. The main preprocessing steps were:

- **Normalization**: All text was converted to a uniform format (e.g. lowercasing, removal of extraneous spaces) to avoid inconsistencies due to capitalization or formatting.
- **Noise filtering**: Non-informative characters and symbols (e.g. stray punctuation, special symbols) that do not contribute to semantic content were removed to reduce noise.
- **Tokenization**: The narratives were segmented into tokens (primarily word-level units), which served as the fundamental elements for subsequent analysis.
- **Lemmatization/stemming**: Words were reduced to their base or root forms (e.g. 'damaging' → 'damage') to collapse inflected variants into a common representation and reduce sparsity.
- **Stop-word refinement**: A domain-specific stop-word list was iteratively developed for maritime accident reports, removing highly frequent but semantically uninformative terms (e.g. function words) while retaining technical phrases relevant to causation.

This preprocessing produced a clean and linguistically normalized corpus suitable for both keyword-based extraction and more contextual, LLM-driven interpretation.

3.2.2. Feature extraction and indicator definition

Next, we transformed the preprocessed text into candidate features and then into the specific indicators required for quantitative modeling. As a supporting step, Bag-of-Words (BoW) and Term Frequency-Inverse Document Frequency (TF—IDF) representations were used to identify salient terms and phrases associated with communication failures, human states, operational stressors, and organizational conditions (Ofori-Boateng et al. 2024).

- **BoW profiling**: Word-frequency patterns highlighted recurring terminology linked to language barriers (e.g. 'misunderstood,' 'no reply,' 'incorrect readback'), human factors (e.g. 'fatigued,' 'overloaded'), and contextual conditions (e.g. 'dense traffic,' 'restricted channel').
- **TF—IDF weighting**: TF—IDF scores emphasized less frequent but diagnostically important phrases, helping to distinguish background narrative from causal cues.
- **Contextual extraction with CustomGPT**: Guided by these profiles and a structured prompt template, the CustomGPT pipeline then extracted case-level values for predefined fields such as language proficiency, communication role violations, stress and fatigue, Cognitive load, navigational complexity, weather, time pressure, crew training, safety culture, and multilingual crew composition. The extraction relied not only on keyword matching but also on sentence-level context to interpret whether a condition was present and at what level.

The result of this step was a set of candidate indicators for each case, aligned with the conceptual domains that would later be modeled as latent constructs in the SEM.

3.2.3. Accident selection and coding

The Only investigation reports in which language and/or communication issues were explicitly identified as causal or contributory were retained, yielding 515 accidents. For these cases, the NLP pipeline was used solely to code the contextual indicators listed in Table 2 (e.g. proficiency, stress, time pressure, safety culture) into their categorical states. In the SEM, Accident Risk is treated as a single latent outcome influenced by these constructs, rather than as a multi-level severity category.

3.2.4. Variable coding and state categorization

After extraction and scenario tagging, the causation-related elements were converted into structured variables with clearly defined categories (states). For each indicator, we specified discrete levels that reflect the intensity or nature of the condition, for example:

- Language proficiency: low, medium, high
- Accent/pronunciation clarity: None, mild, severe
- Stress and fatigue: None, mild, severe
- Cognitive load: low, medium, high
- Navigational complexity: low, medium, high
- Weather: Clear, Fog, Storm, rough sea
- Time pressure: None, moderate, high
- Crew training: poor, average, good
- Safety culture: strong, moderate, weak
- Multilingual crew composition: Single/monolingual, Bilingual, multilingual

States were determined by combining textual cues in the reports with expert-informed coding rules. Where wording was ambiguous, two of the authors jointly reviewed the relevant excerpts and agreed on the final classification. This procedure ensured that the categorical codings remained faithful to the narrative evidence while being suitable for subsequent categorical factor analysis and SEM.

3.2.5. Structured dataset for factor analysis and SEM

Once coding was complete, all case-level variables were assembled into a structured dataset. Each accident report corresponded to a single row, and each indicator (linguistic, human, operational, organizational) was represented as a categorical column. The resulting dataset contained 515 fully coded reports, after excluding cases with incomplete or inconsistent information. This was saved as a final CSV file.

This structured dataset served three purposes:

(1) **Exploratory factor analysis (EFA)**: EFA on the polychoric/tetrachoric correlation matrix was used to confirm the latent structure (communication barrier, human factors, operational context, organizational factors).

(2) **Confirmatory/structural modeling**: The same indicators were then used in a covariance-based SEM with a WLSMV estimator, which is appropriate for ordered categorical indicators and models thresholds explicitly.
(3) **Robustness checks**: Sensitivity analyses (e.g. comparing dummy-coded versus original ordinal specifications) were conducted to verify that the latent structure was not an artefact of the coding scheme.

By moving systematically from unstructured narrative reports to a rigorously structured set of categorical indicators, the NLP pipeline provided a transparent bridge between qualitative accident evidence and the quantitative SEM used to model accident risk.

We verified extraction quality with a random 30% manual audit of reports, checking each extracted variable against the study's operational definitions and inclusion criteria. The audit found no systematic discrepancies; occasional ambiguous cases were adjudicated, and the inclusion criteria were clarified accordingly. The finalization of variables and corresponding states in the structured data was conducted in concurrence with the opinion of subject matter experts. The consulted experts formed a multidisciplinary panel comprising a ship captain, a senior safety engineer, a vessel traffic service (VTS) officer, a professor of maritime safety, and a professor of applied linguistics (Maritime English). Together, they represent ship operations and management, safety engineering and regulation, traffic control, and maritime English/linguistics, providing both practical and academic perspectives on communication-related accident causation. The details of these experts are provided below in Table 1.

To ensure that the variable set reflected both the literature and operational realities, we used a structured, workshop-based consensus process. First, the authors compiled an evidence pack from prior studies and narrative accident reports (definitions, inclusion/exclusion criteria, and illustrative excerpts). Next, a multi-disciplinary panel of maritime experts, as defined, independently pre-rated each candidate variable on two dimensions: relevance to communication-related accident causation and codability/observability in investigation narratives.

We then held facilitated workshops. For each variable, the moderator presented summaries of the pre-ratings and representative report excerpts. Divergent views were reconciled via structured discussion anchored to the inclusion criteria (relevance, distinctness/non-redundancy with other variables, and codability). Where disagreements persisted, we refined the operational definition and, in cases, merged overlapping constructs to maintain parsimony without losing substantive content. A brief post-discussion re-rating confirmed consensus before finalization.

Table 1. Profile description of the consulted experts.

Expert ID	Role/Domain	Key Qualifications/Certifications	Experience (approx.)
E1	Ship Captain	Master's Degree	30 Years
E2	Senior Safety Engineer	Master's Degree	22 Years
E3	Vessel Traffic Service (VTS) Officer	Master's Degree	18 Years
E4	Professor of Maritime Safety	Ph.D.	16 Years
E5	Professor of Applied Linguistics (Maritime English)	Ph.D.	10 Years

Table 2. Initial variables (ordinal categories and descriptions).

Variable	Coding	Description
Communication Role Violation	Yes = 1 No = 0	Violations of the team roles and rules regarding communication and coordination.
Language Proficiency	Low = 0 Medium = 1 High = 2	Lack of language proficiency
Accent/ Pronunciation	None = 0 Mild = 1 Severe = 2	Unclear pronunciation
Experience	Low = 0 Medium = 1 High = 2	Work experience with multinational crews and standard communication protocols
Stress and Fatigue	None = 0 Mild = 1 Severe = 2	Stress and fatigue effect on cognitive function and language comprehension
Cognitive Load	Low = 0 Medium = 1 High = 2	High workload leading to errors in language processing
Navigational Complexity	Low = 0 Medium = 1 High = 2	High traffic and narrow passages.
Weather Conditions	Clear = 0 Fog = 1 Storm = 2 RoughSea = 3	Impact of weather, visibility, and sea conditions on communication needs.
Time Pressure	None = 0 Moderate = 1 High = 2	Criticality of the situation and pressure to make quick decisions, impacting communication clarity
Crew Training	Poor = 0 Average = 1 Good = 2	Training on standardized communication practices
Safety Culture	Weak = 0 Moderate = 1 Strong = 2	Emphasis on closed-loop communication and standard language implementation and regular language proficiency assessments
Multilingual Crew Composition	Single = 0 Bilingual = 1 Multilingual = 2	The presence and speaking of multiple native languages among crew members

Note. Numeric codes are ordered category labels rather than numeric scores. All indicators were treated as categorical variables: exploratory factor analysis was based on polychoric/tetrachoric correlations, and the SEM used a categorical estimator with thresholds. No interval meaning is assumed for the codes (0/1/2/3), which simply index ordered response levels.

Finally, the variables and coding details extracted from the detailed accident reports are depicted in Table 2. The number of accident reports in our sample size is in concurrence with relevant literature recommending at least ten observations for each variable (Austin and Steyerberg 2015; Sheykhfard et al. 2020).

3.3. SEM

This study adopts a covariance-based SEM framework to model simultaneously (i) the relationships between observed indicators and their latent constructs (measurement model), and (ii) the relationships among latent constructs (structural model). This separation enables explicit treatment of measurement error and formal testing of inter-dependencies among constructs relevant to maritime safety. A typical SEM has two parts: the measurement and the structural components. The measurement model links observed indicators to latent variables; the structural model specifies relations among latent variables (Sheykhfard and Haghighi 2020). In SEM, models comprise observed indicators and latent constructs. Latent variables are unobserved constructs inferred from multiple indicators; their measurement is obtained by estimating factor loadings and thresholds that link indicators to the construct while accounting for measurement error.

The measurement part can be mathematically expressed as;

$$y_i = \lambda_{\mathrm{i}}\eta + \varepsilon_{\mathrm{i}} \tag{1}$$

where,

y_i: Observed indicator (e.g. survey items or measurements)

η: Latent variable (construct being measured)

λ_{i}: Factor loading, representing the strength of the relationship between the latent variable and its indicator

ε_{i}: Measurement error for indicator y_i

For scenarios with multiple latent variables and multiple indicators, the mathematical expression turns into;

$$Y = \Lambda_\eta \eta + \varepsilon \tag{2}$$

Where,

Y: Vector of observed indicators

Λ_η: Matrix of factor loadings

η: Vector of latent variables

ε: Vector of measurement errors

Because our indicators are binary/ordinal, the measurement model is implemented via a latent-response (threshold) formulation:

$$y^*{}_i = \lambda_{\mathrm{i}}\eta + \varepsilon_{\mathrm{i}}, \quad y_i = c if \tau_{c-1} < y^*{}_i \leq \tau_c \tag{3}$$

where $y^*{}_i$ is a continuous latent response, and τ_c are thresholds partitioning categories (e.g. Low/Medium/High). In this formulation, the observed categorical responses arise by discretizing an underlying continuous propensity; estimation therefore relies on polychoric/tetrachoric associations in factor analysis and on a categorical SEM estimator that explicitly models thresholds in the measurement model.

While the structural component can be mathematically represented as;

$$\eta = \beta\eta + \Gamma\xi + \zeta \tag{4}$$

Where,

η: Endogenous latent variables (dependent constructs)

ξ: Exogenous latent variables (independent constructs)

β: Matrix of path coefficients between endogenous latent variables

Γ: Matrix of path coefficients from exogenous to endogenous latent variables

ζ: Error term for the structural relationships (residuals)

Estimation: Exploratory factor analysis was conducted on an asymptotic polychoric/tetrachoric correlation matrix (tetrachoric for binary indicators; polychoric for ordinal indicators) with varimax rotation to identify the latent structure. The subsequent confirmatory factor analysis and structural models were estimated using a categorical covariance-based SEM estimator, specifically the robust weighted least squares estimator with mean and variance adjustment (WLSMV), with a probit link, threshold parameters for all categorical indicators, and robust (mean/variance-adjusted) χ^2 and standard errors.

Identification and scaling: Latent variable scales were set via the marker-variable method (one loading per factor fixed to 1.0); remaining loadings and thresholds were freely estimated.

Model evaluation: Model evaluation relied on standard covariance-based SEM fit indices for categorical data: χ^2 /df, RMSEA, SRMR, CFI, and TLI, supplemented by additional indices (e.g. GFI, AGFI, NFI, IFI) reported in the Results to provide a comprehensive assessment of global fit. Where AIC/BIC are referenced, they stem from a secondary maximum-likelihood sensitivity run used only for parsimony comparison; all substantive inferences are based on the categorical WLSMV estimation.

4. Results and discussion

4.1. Factor analysis

Exploratory factor analysis (EFA) was conducted to identify the latent structure underlying the observed indicators prior to SEM. For multi-category indicators in Table 2, we adopted a two-step strategy. In the primary EFA, selected categorical variables were represented by a small set of binary ('dummy') indicators to isolate the contribution of specific high-risk states (e.g. low proficiency vs. all other proficiency levels). For example, three indicators were created for language proficiency: (1) Low vs. other levels, (2) Medium vs. other levels, and (3) High vs. other levels (Gaweesh Sherif, Ahmed Irfan, and Ahmed Mohamed 2024; Sheykhfard and Haghighi 2020; Sheykhfard et al. 2025). This allowed us to test directly whether particular categories (such as low proficiency or severe fatigue) clustered together in the same latent factor. Although broader language frameworks (e.g. CEFR—Common European Framework of Reference for Languages; ICAO LPRs—International Civil Aviation Organization Language Proficiency Requirements) specify six levels, they are commonly summarized into three bands (beginner/intermediate/proficient) (Mönnigmann and Čulić-Viskota 2017). In maritime practice, there is no universally adopted multi-level standard (despite tests such as the ISF, International Shipping Federation, Marlins English Test, and initiatives like SeaTalk) (Mönnigmann and Čulić-Viskota 2017), and accident narratives provide only coarse descriptors; therefore, we used a three-tier scheme, operationalized via ordered categorical codes and, in the primary EFA, by corresponding dummy indicators for interpretative clarity.

EFA was then performed on an asymptotic polychoric/tetrachoric correlation matrix (tetrachoric for binary indicators; polychoric for ordinal indicators), which is appropriate for categorical data. To assess robustness and address potential concerns about dummy-coding, we also ran a sensitivity EFA using the original ordinal codings (without binary expansion) on the same polychoric/tetrachoric correlation matrix. Both specifications produced the same four-factor structure and very similar loading patterns, with the same indicators showing salient loadings on the same latent constructs; the dummy-coded solution is reported in Table 4. The next step was to conduct factor analysis employing the varimax rotation. In this regard, first, the Kaiser-Meyer-Olkin (KMO) and Bartlett's test of sphericity were conducted to evaluate the suitability for factor analysis. The KMO values of 0.83

and Bartlett's p-value of 0.01 indicate good suitability for factor analysis. Varimax rotation is a widely used technique designed to simplify the interpretation of factors by increasing the larger loadings and reducing the smaller loadings (Lee, Chung, and Son 2008).

This study used varimax to enhance simple structure, i.e. to increase large loadings and suppress small cross-loadings, thereby improving the interpretability of the factor pattern for categorical indicators. The loadings of these variables were checked for four latent constructs, including Human Factor, Communication Barrier, Operational Context, and Organizational Factors. As aligned with the aim of this study, the purpose of these latent variables was to determine the impact of human aspects, operational variability, communication hindrances, and organizational dimensions on the efficient language and communication of the ship crew and the causation of accidents because of the failures in closed-loop communication. In this regard, loadings above 0.6 were considered high, while those below 0.4 were considered low. These thresholds prioritize substantive salience over mere statistical significance, given our sample size ($n = 515$), and align with established guidance (Comrey and Lee 2013; Hair 2009; Hair et al. 2019; Stevens 2002). Across both the dummy-coded and ordinal-coded EFAs, the same indicators exceeded these thresholds on the same factors, supporting the stability of the latent structure. Table 3 reports indicators with salient loadings (as defined above) on one of the four latent constructs.

As shown in Table 3, Factor 1 (Communication Barrier) is characterized by Communication Role Violation (Yes), Language Proficiency (Low), and Accent/Pronunciation (Severe). Factor 2 (Human Factors) is marked by Experience (Low), Stress and Fatigue (Severe), and Cognitive Load (High). Factor 3 (Operational Context) is defined by Navigational Complexity (High), Weather Conditions (Rough sea), and Time Pressure (High). Factor 4 (Organizational Factors) is defined by Crew Training (Poor), Safety Culture (Weak), and Multilingual Crew Composition (Multilingual). Indicators without salient loadings are omitted for parsimony. This factor structure informed the specification of the measurement model in the subsequent SEM.

4.1.1. Sensitivity analysis of factor structure

To verify that the factor solution is not an artefact of the dummy-coding scheme, we conducted a robustness EFA using the original ordinal coding of the indicators (without

Table 3. Factor analysis results of varimax rotation.

Variable	Factor 1	Factor 2	Factor 3	Factor 4
Communication Role Violation (Yes)	0.879	0.051	0.007	0.007
Language Proficiency (Low)	0.831	0.032	−0.028	0.042
Accent/Pronunciation (Severe)	0.726	0.147	−0.071	0.298
Experience (Low)	0.195	0.714	0.067	0.061
Stress and Fatigue (Severe)	0.062	0.772	0.198	0.011
Cognitive Load (High)	0.194	0.893	0.072	0.187
Navigational Complexity (High)	0.002	−0.033	0.906	0.075
Weather Conditions (RoughSea)	0.021	−0.011	0.976	−0.078
Time Pressure (High)	−0.006	0.034	0.927	−0.013
Crew Training (Poor)	−0.019	0.076	0.179	0.827
Safety Culture (Weak)	0.007	0.298	0.057	0.945
Multilingual Crew Composition (Multilingual)	0.011	0.307	0.067	0.924

Note. Loadings are standardized estimates from polychoric/tetrachoric EFA with varimax rotation.

Table 4. Indicator specification for robustness EFA using original ordinal codings.

Construct	Indicator	Code	Category/meaning	Scale type
Communication Barrier	Communication Role Violation	0	No	Binary
		1	Yes	Binary
	Language proficiency	0	Low	Ordinal
		1	Medium	Ordinal
		2	High	Ordinal
	Accent/pronunciation	0	None	Ordinal
		1	Mild	Ordinal
		2	Severe	Ordinal
Human Factors	Experience	0	Low	Ordinal
		1	Medium	Ordinal
		2	High	Ordinal
	Stress and fatigue	0	None	Ordinal
		1	Mild	Ordinal
		2	Severe	Ordinal
	Cognitive load	0	Low	Ordinal
		1	Medium	Ordinal
		2	High	Ordinal
Operational Context	Navigational complexity	0	Low	Ordinal
		1	Medium	Ordinal
		2	High	Ordinal
	Weather conditions	0	Clear	Ordinal
		1	Fog	Ordinal
		2	Storm	Ordinal
		3	Rough sea	Ordinal
	Time pressure	0	None	Ordinal
		1	Moderate	Ordinal
		2	High	Ordinal
Organizational Factors	Crew training	0	Poor	Ordinal
		1	Average	Ordinal
		2	Good	Ordinal
	Safety culture	0	Weak	Ordinal
		1	Moderate	Ordinal
		2	Strong	Ordinal
	Multilingual crew composition	0	Single language	Ordinal
		1	Bilingual	Ordinal
		2	Multilingual	Ordinal

Table 5. Robustness EFA (polychoric) factor loadings for original ordinal indicators.

Indicator	Factor 1	Factor 2	Factor 3	Factor 4
Language proficiency (0–2)	0.79	0.018	0.05	0.09
Accent/pronunciation clarity (0–2)	0.75	0.021	0.02	0.20
Communication role violation (0/1)	0.83	0.12	0.08	0.11
Experience (0–2)	0.12	0.71	0.10	0.06
Stress and fatigue (0–2)	0.14	0.80	0.15	0.04
Cognitive load (0–2)	0.20	0.83	0.12	0.18
Navigational complexity (0–2)	0.05	0.08	0.86	0.10
Weather conditions (0–2)	0.02	0.03	0.88	0.06
Time pressure (0–2)	−0.09	0.19	0.82	0.07
Crew training (0–2)	−0.10	0.15	0.12	0.79
Safety culture (0–2)	0.11	0.21	0.10	0.85
Multilingual crew composition (0–2)	0.14	0.16	0.14	0.81

binary expansion). As in the main analysis, this EFA was based on a polychoric correlation matrix with varimax rotation, and the four-factor solution was retained. The indicator specification for this robustness check is summarized in Table 4, and the resulting loadings are reported in Table 5.

The alternative specification recovers the same four latent constructs and the same set of indicators with salient loadings on each factor: variables associated with communication barriers, human factors, operational context, and organizational conditions continue to cluster on their respective factors, and no indicator changes its primary factor. While some loading magnitudes shift slightly compared with the dummy-coded solution, the pattern of high versus low loadings remains substantively unchanged. This confirms that the four-factor structure is driven by the underlying accident patterns rather than by the specific coding scheme, and it supports the use of the more interpretable dummy-coded solution reported in the main EFA for subsequent SEM.

4.2. SEM analysis and discussion

The finalized SEM model for the language and communication issues induced accidents under the influence of human, operational, and organizational factors is depicted in Figure 2. In Figure 2, values on paths are standardized (STDYX) WLSMV estimates; parentheses report robust SE and z-statistics. Because results are fully standardized, coefficients are directly comparable across indicators and constructs. Within the Human Factors construct, Stress and Fatigue (Severe) ($\lambda = 0.88$, SE = 0.05, $z = 17.6$) and Cognitive Load (High) ($\lambda = 0.85$, SE = 0.05, $z = 17.0$) load strongly, with Experience (Low) ($\lambda = 0.69$, SE = 0.07, $z = 9.86$) also salient. Higher levels of stress and fatigue are reported to have a major role in accident causation (Albert 2024; Setiawan and Razak Abdul Hadi 2025), as compared to a low experience. Likewise, the cognitive load on the involved personnel has a higher impact. This also compromises the role of humans in ensuring efficient closed-loop communication and safer navigation (M. H. Lützhöft and Dekker 2002; M. Lützhöft, Grech, and Porathe 2011). These indicators reflect conditions that are known to impair communication and decision quality on the bridge. At the structural level, Human Factors shows a large standardized association with Accident Risk ($\beta = 0.60$, SE = 0.05, $z = 12$), indicating that higher human-state strain is strongly linked with increased accident risk. This pattern is consistent with maritime safety evidence that human-state stressors are central contributors to accidents, particularly where multilingual communication is required. Practically, the findings motivate interventions that manage workload and fatigue (e.g. watch scheduling, task redistribution, and fatigue risk management), which are likely to improve communication reliability and decision performance on the bridge.

Moving towards the latent construct of the communication barrier, language proficiency appears as the most significant and critical factor with a high impact on the language barrier in communication. While communication issues are considered a significant contributor to the maritime accidents in the contemporary literature (Boström 2020b, 2021; L. Liu et al. Yin et al. 2024, 2025). In the standardized (STDYX) solution, Language Proficiency (Low) ($\lambda = 0.91$, SE = 0.03, $z = 30.33$) and Communication Role Violation (Yes) ($\lambda = 0.89$, SE = 0.04, $z = 22.2$) show the strongest loadings on the Communication Barrier latent, with Accent/Pronunciation (Severe) also prominent; together these indicators capture decoding/encoding difficulty, adherence to closed-loop protocols, and intelligibility. At the structural level, the Communication Barrier latent exhibits a moderate—large, statistically significant association with Accident Risk ($\beta = 0.55$; SE = 0.05 and $z = 11$), indicating that deficits in proficiency

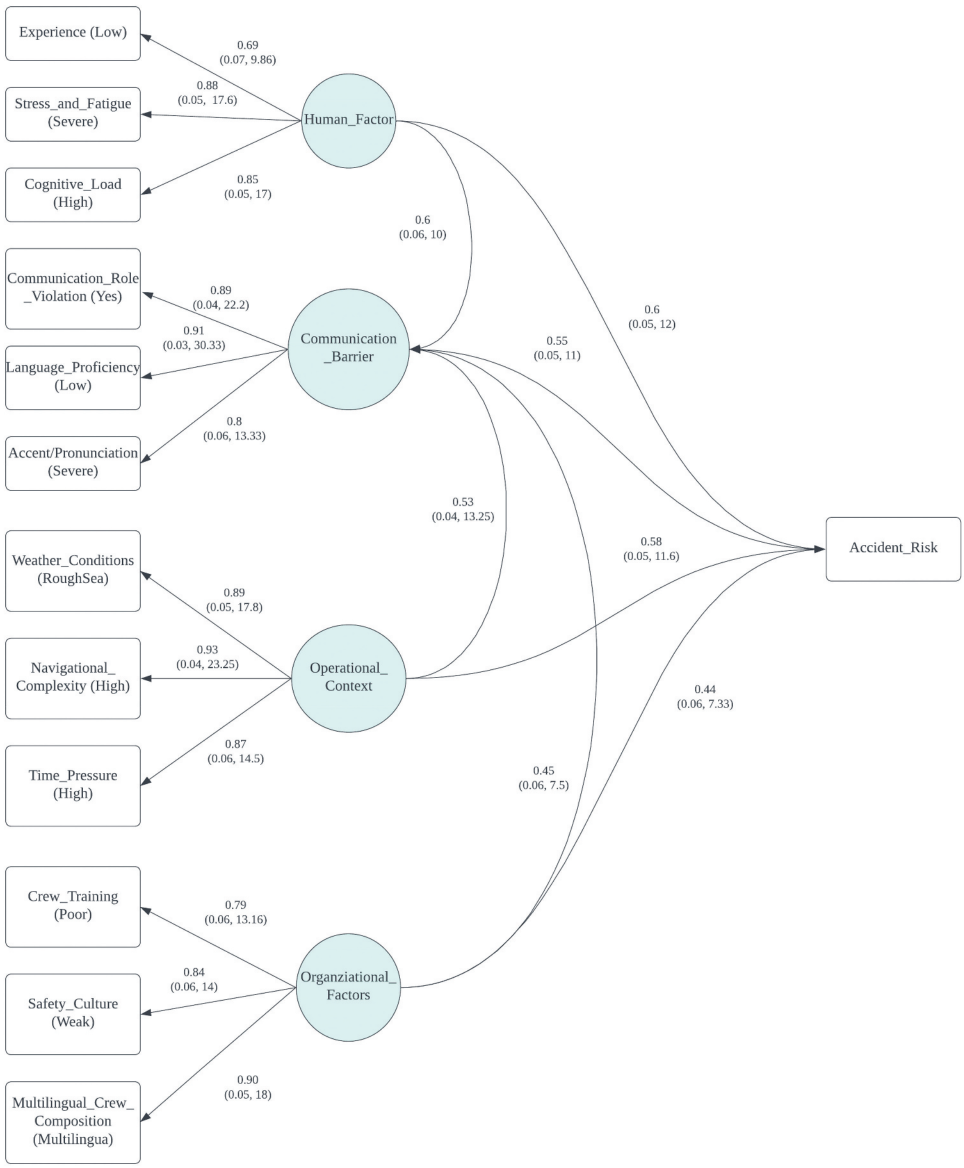

Figure 2. Finalized SEM (standardized WLSMV estimates with se and z) model of the language barriers in maritime safety with performance and fit indices of $R^2 = 0.845$ for accident risk, RMSEA = 0.024, SRMR = 0.012, CFI = 0.962, and TLI = 0.971.

and breakdowns in role-based communication are strongly linked to higher accident likelihood. Though the role of accent and pronunciation is also significant and notable in this regard, however, the language efficiency of the SMCP is revealed as a high-stakes factor. Mechanistically, lower proficiency increases the cognitive effort required to encode/interpret SMCP and non-routine utterances, accent/phonology reduces intelligibility under noise, and both raise the probability of miscommunication (missed read-backs, incomplete confirmations). This factor is more critical on international voyages with multilingual crews and international exposures. These risks are amplified under operational stressors (time pressure, rough sea) when crews are prone to omission, truncation, or code-switching; the moderate latent correlation between Communication Barrier and Operational Context in our model is consistent with this amplification pathway. A sound language proficiency and adherence to the SMCP can significantly avert the risk of miscommunication and hence the occurrence of undesired evidence. This aspect becomes further critical in distress situations where the involved personnel are under pressure, facing the challenges of time, navigational complexity, and rough seas, where timely and accurate communication and decision-making are of critical essence. Practically, this supports targeted SMCP training with closed-loop drills, accent-aware radio discipline, and pre-briefed standard phraseology for high-risk evolutions (pilotage, collision-avoidance) to reduce violation risk and stabilize team coordination.

Continuing the discussion on the impact of surrounding conditions, the latent variable of operational context is considered, for which the impact of observed variables of weather conditions, navigational complexity, and time pressure is measured. In the standardized measurement model, Navigational Complexity (High) ($\lambda = 0.93$, $SE = 0.04$, $z = 23.25$), Weather Conditions (Rough sea) ($\lambda = 0.89$, $SE = 0.05$, $z = 17.8$), and Time Pressure (High) ($\lambda = 0.87$, $SE = 0.06$, $z = 14.5$) load strongly on the Operational Context latent, reflecting environments that increase message volume, urgency, and channel noise. Navigational complexity substantially heightens language and communication barriers by requiring precise and timely exchanges of information in high-pressure scenarios, such as navigating through narrow straits, heavily trafficked sea lanes, or adverse weather conditions. Practical challenges arise when crew members, often from multilingual backgrounds, must interpret navigation commands involving technical jargon or coordinate complex maneuvers like course alterations and collision avoidance. Inadequate understanding due to accents, pronunciation differences, or ambiguous phrasing can delay decision-making or lead to incorrect actions. This lack of clarity increases the likelihood of near-misses or collisions, especially in situations demanding rapid responses, thereby elevating accident-causation risks associated with language and communication gaps in maritime transportation Structurally, Operational Context shows a significant, positive standardized association with Accident Risk ($\beta = 0.58$; $SE = 0.05$ and $z = 11.6$), consistent with the notion that complex, time-critical, and weather-degraded settings increase failure opportunities in communication and decision pipelines (Yin et al. 2025). Time pressure is particularly critical during near-miss scenarios or maritime emergencies, where rapid and precise communication is essential to prevent escalation. Under such conditions, the urgency to act often leads to rushed exchanges, increasing the likelihood of misinterpretations or incomplete messages, especially among multilingual crews. For example, during collision avoidance maneuvers or emergency evacuation procedures, time pressure can compromise adherence to standard communication protocols, such as repeating instructions or confirming understanding. This heightens the risk of errors, delays in response, and coordination failures, directly amplifying the accident causation potential in already high-stress situations. The observed interplay in our model, moderate correlations between Operational Context and both Communication Barrier, supports a mechanism in which context elevates cognitive load and compresses time for closed-loop verification, thereby exacerbating language-related breakdowns. Operationally, this points to 'sterile bridge' protocols during critical phases, pre-briefed communication checklists, explicit read-back/confirm practices under time pressure, and crew pairing that ensures at least one high-proficiency operator controls external comms in congested/rough-weather transits.

Another important aspect of the current model is the impact of organizational factors, for which the latent variables of crew training, safety culture, and multilingual crew composition were considered. In the measurement model, Crew Training (Poor) ($\lambda = 0.79$, SE = 0.06, z = 13.16), Safety Culture (Weak) ($\lambda = 0.84$, SE = 0.06, z = 14), and Multilingual Crew Composition (Multilingual) ($\lambda = 0.90$, SE = 0.05, z = 18) load strongly on the Organizational Factors latent (see Figure 2), indicating that both enabling conditions (training, culture) and crew language mix are integral to the organizational substrate of communication. Organizational factors, including crew training, safety culture, and multilingual crew composition, significantly shape how language and communication work in practice. Inadequate crew training can lead to a poor understanding of communication protocols, resulting in misinterpretations and failures in critical instructions, while a weak safety culture undermines the importance of clear, precise communication, especially in high-risk environments. In this study, a higher score on Multilingual Crew Composition does not mean that multilingual crews are inherently unsafe; rather, it captures situations in which several language groups are present on board without adequate organizational mitigation (e.g. a clearly defined working language, robust SMCP training, or structured use of bilingual 'bridging' personnel). In such under-prepared settings, linguistic diversity increases coordination demands and makes the system more sensitive to deficiencies in proficiency, accent intelligibility, and adherence to closed-loop protocols. Structurally, Organizational Factors shows a small-to-moderate positive association with Accident Risk ($\beta = 0.44$, SE = 0.06, z = 7.33), consistent with an upstream, enabling role: training and culture primarily buffer violation likelihood, whereas poorly managed multilingual composition amplifies communication demands and exposes latent weaknesses in procedures. Practically, this pattern suggests two high-leverage organizational levers. First, institutionalize closed-loop norms and SMCP adherence through recurrent, scenario-based drills, pre-sail briefs, and supervisory reinforcement (audit read-backs, require confirmations for high-risk commands). Second, design watch compositions and role assignments to ensure language redundancy and clarity of radio responsibility, for example, assign external communications to a demonstrably high-proficiency operator in congested or weather-degraded transits, pair bilingual officers with less proficient crewmates for internal coordination, and schedule overlaps that maintain a minimum language capability on watch. These measures target the organizational substrate that, in our model, covaries with barriers and context, thereby reducing the probability that routine strain escalates into communication breakdowns.

4.3. Model Fit Assessment

This study assesses statistical significance using robust z-statistics from the WLSMV estimator. Absolute $z \geq 1.96$ indicates statistical significance at the 95% confidence level. Z-statistics quantify precision (via SE) and significance; effect magnitudes are conveyed by the standardized (STDYX) coefficients. In the finalized model (Figure 2), the core structural paths and salient loadings achieve $|z| \geq 1.96$, indicating statistically reliable associations. The R^2 value, another key metric, quantifies the proportion of variation in the endogenous variables explained by the exogenous variables. For the endogenous latent Accident Risk, the R^2 is 0.845, indicating that the latent predictors account for ~84.5% of variance, substantively high and consistent with the overall fit indices.

The chi-square ($\chi 2$) statistic is used to assess how closely the implied variance-covariance matrices align with the sample variance-covariance matrices. A chi-square value of zero indicates perfect alignment, while an increasing value reflects a greater discrepancy between the implied and observed matrices. In model evaluation, smaller or non-significant $\chi 2$ values are preferred. However, reliance solely on $\chi 2$ has been criticized, particularly in cases with sample sizes exceeding 200 (Lomax 2004). In this study, we report the *robust* (mean—variance adjusted) χ^2 associated with the WLSMV estimator; the chi-square test may falsely suggest significant deviations, which is why the ratio of $\chi 2$ to degrees of freedom (df) is often considered a more reliable fit index. Researchers such as Marsh and Hocevar recommend a $\chi 2$: df ratio between 2:1 and 5:1 (Marsh and Hocevar 1988).

Root Mean Square Error of Approximation (RMSEA) is another widely used criterion for model fit, with values under 0.05 signaling a well-fitting model (Byrne 2013). The Standardized Root Mean Square Residual (SRMR) complements RMSEA by summarizing average standardized residuals; values < 0.08 (often < 0.05) indicate close fit. Goodness-of-Fit Index (GFI) compares the variance-covariance in the sample to that predicted by the estimated model, with values between 0.9 and 1.0 indicating a strong fit (Lomax 2004). The Residual Mean Square (RMR) indicates the closeness between the reproduced and observed covariance matrices, with acceptable thresholds depending on the study's objectives and the correlation values (Bagozzi and Yi 1989). The Normed Fit Index (NFI) evaluates the fit by comparing a restricted model to a full model, using a baseline null model (Lomax 2004). Comparative Fit Index (CFI) and Incremental Fit Index (IFI), derived from the NFI, account for sample size, with both indices ranging from 0 (indicating no fit) to 1 (indicating perfect fit). In line with CB-SEM practice, we emphasize χ^2 /df, RMSEA, SRMR, CFI, and TLI as primary evidence of fit, with GFI/AGFI/NFI/RMR reported as supplementary.

In the present model, the robust (WLSMV) chi-square statistic was significant with a value of 118.43, degrees of freedom (df) = 82, and a $p < .001$. Despite the significant $\chi 2$, the $\chi 2$ to df ratio was below 2, confirming a good model fit according to Ullman (Ullman and Bentler 2012). Both the RMSEA values were under the accepted threshold of 0.05, and SRMR was 0.012 (close fit), while the Adjusted Goodness-of-Fit Index (AGFI) and GFI exceeded the 0.90 standard. The relative fit indices, IFI and CFI, were both greater than 0.90. The NFI was also above the ideal threshold of 0.90, showcasing the aspects of a very good fit. For parsimony comparison, information criteria from a maximum-likelihood sensitivity run were favorable (AIC = 98; BIC = 122); note that AIC/BIC are

Table 6. Summary of the model goodness-of-fit indices.

Parameter	Value	Standard
χ^2 / df	1.4442	<2
RMSEA	0.024	<.05
SRMR	0.012	<.08
CFI	0.962	≥.90
Tucker–Lewis Index (TLI)	0.971	≥.90
PGFI	0.643	≥.50 (parsimony)
PNFI	0.651	≥.50 (parsimony)
AIC†	98	Lower vs alternatives
BIC†	122	Lower vs alternatives
Supplementary (legacy) indices		
GFI	0.985	≥.90
AGFI	0.977	≥.90
NFI	0.946	≥.90
RMR	0.00073	<.05
IFI	0.952	≥.90

†AIC/BIC computed from a maximum-likelihood sensitivity fit; AIC/BIC are not defined for WLSMV and are reported only for parsimony comparison.

not defined for WLSMV and are reported here solely for comparative context. A detailed summary of the goodness-of-fit indices for the model path diagram can be found in Table 6.

5. Practical implications

Grounded in the empirical patterns of our SEM analysis, the following recommendations translate specific path coefficients and loadings into targeted controls for bridge communication, human-state management, context-triggered procedures, and organizational governance within the Safety Management System (SMS).

Communication assurance for high-risk operations (driven by Communication Barrier → Accident Risk, $\beta = 0.55$; and $\lambda = 0.91$ / 0.89 for Low Proficiency and Role Violation).

Given that the Communication Barrier latent shows a moderate—large association with Accident Risk, and that Language Proficiency (Low) and Communication Role Violation carry the highest loadings on this factor, interventions should prioritize preventing miscommunication exactly where its consequences are most severe. In practice, this means: (i) establishing a minimum SMCP/Maritime English proficiency band for key bridge and communication roles; (ii) providing targeted remediation or mentoring for crew identified as low proficiency; and (iii) in clearly defined high-risk phases (pilotage, restricted waters, heavy traffic, rough weather), activating a heightened communication mode: SMCP-dominant phraseology; one-order-one-acknowledgement with mandatory read-back; numerals spoken digit-by-digit; and a deliberate speaking rate on VHF/bridge. To address feasibility, this mode is not intended for continuous use but for short, risk-triggered windows anchored in existing checklists (e.g. pilot boarding, narrow-channel entry), supported by simple tools such as bridge-side phrase cards and brief ‘verbalization drills’ at watch handover. This risk-based, time-limited activation balances operational workload with the need to reduce communication role violations when the model indicates they matter most.

Managing human state in real time (driven by Human Factors → Accident Risk, β = 0.60; and λ = 0.88 / 0.85 / 0.69 for Stress & Fatigue, Cognitive Load, Experience).

Human Factors has the largest standardized path to Accident Risk in the model, dominated by high loadings for Stress and Fatigue and Cognitive Load. This suggests that even well-designed communication protocols can fail if the human state is degraded. Practically, operators can (i) embed fatigue risk management into watch schedules (protecting minimum rest, avoiding repeated 'heavy traffic + night watch' assignments), (ii) apply explicit workload management on the bridge (clearly assigning who monitors traffic, who manages comms, who handles documentation/alarms), and (iii) require short pre-briefs before expected high-demand legs to fix intentions, call-outs, and contingency wording. Simulator training can mirror the SEM findings by combining high workload, time pressure, and multilingual crews, with debriefs focused on phraseology, confirmations, and hand-offs. These measures directly target the high-loading human-state indicators that the SEM links to Accident Risk.

Context-triggered procedures and team composition (driven by Operational Context → Accident Risk, β = 0.58; and λ = 0.93 / 0.89 / 0.87 for Navigational Complexity, Weather, Time Pressure).

Because Operational Context shows a strong positive association with Accident Risk, with navigational complexity, rough seas, and time pressure loading heavily on this latent, procedures should be explicitly context-sensitive. Companies can define objective triggers, traffic density, visibility, sea state, and proximity to narrow channels, for switching into the heightened communication mode described above and for adjusting task allocation (e.g. dedicating one officer to communication only). Watch composition can be planned so that at least one high-proficiency speaker is always present during these high-demand segments, with a less proficient crew paired with that person for critical communications. In this way, the heightened mode becomes a feasible, event-driven layer added on top of routine operations rather than an unrealistic expectation for continuous, strict SMCP use.

Organizational embedding and monitoring (driven by Organizational Factors → Accident Risk, β = 0.44; and λ = 0.79 / 0.84 / 0.90 for Training, Safety Culture, Multilingual Composition).

Organizational Factors show a smaller but still significant path to Accident Risk, consistent with an upstream enabling/buffering role. Crew Training and Safety Culture load strongly on this latent alongside Multilingual Crew Composition, indicating that mixed-language crews become risky primarily when they are not supported by robust organizational practices. Accordingly, operators should: (i) incorporate SMCP/closed-loop communication (including multilingual-crew scenarios) into initial and recurrent training; (ii) encode language-aware watch composition and role allocation into the SMS (e.g. assigning external VHF to a demonstrably high-proficiency officer, using bilingual 'bridging' personnel where needed); and (iii) monitor communication-related performance through simple indicators—such as the proportion of critical orders with read-back, distribution of proficiency levels per watch, and logged Communication Role Violations or near-misses involving miscommunication. Feeding these indicators back into training, crew assignment, and procedure updates creates a learning loop that reflects the SEM finding that organizational conditions shape how language and human-state factors translate into accident risk.

Overall, these recommendations are direct operationalizations of the estimated paths and loadings in the SEM: communication controls are prioritized because the Communication Barrier latent (driven by low proficiency and violations) strongly predicts Accident Risk; human-state management is emphasized because Human Factors has the largest structural effect; context-triggered protocols arise from the strong Operational Context path; and organizational embedding is justified by the significant, enabling role of Organizational Factors.

6. Conclusion

This study addresses a critical yet underexplored domain in maritime safety: the role of language and communication failures in accident causation. Given that human factors contribute to 80–90% of maritime accidents, and communication barriers often serve as a pivotal catalyst, this investigation was imperative to identify and quantify the underlying mechanisms driving such incidents. The scarcity of structured data on language and communication-related factors further underscores the necessity of this research. By leveraging advanced NLP techniques and SEM, this study not only fills a significant gap in the literature but also provides actionable insights for enhancing maritime safety. The SEM framework was particularly suitable for this investigation due to its ability to simultaneously analyze complex, multi-variable relationships, allowing us to disentangle the intricate interplay between human, operational, and organizational factors.

The findings reveal that human factors ($\beta = 0.60$, SE = 0.05, z = 12), particularly stress and fatigue ($\lambda = 0.88$, SE = 0.05, z = 17.6), exert the largest standardized effects on Accident Risk, followed by Communication Barrier ($\beta = 0.55$; SE = 0.05 and z = 11) (driven primarily by Language Proficiency ($\lambda = 0.91$, SE = 0.03, z = 30.33), Communication Role Violation, and Accent/Pronunciation) (see Figure 2). Operational Context (Navigational Complexity, Weather, Time Pressure) further intensifies these risks by elevating processing demand and compressing communication windows, while Organizational Factors (Crew Training, Safety Culture, Multilingual Crew Composition) show a smaller but significant association, consistent with an upstream enabling/buffering role. Overall model adequacy is strong, robust (WLSMV) χ^2 (82) = 118.43, $p < .001$; χ^2 /df = 1.44; RMSEA = 0.024; SRMR = 0.012; CFI = 0.962; TLI = 0.971, and the endogenous latent Accident Risk is well explained ($R^2 = 0.845$).

The practical applications of these findings are manifold. Maritime organizations can leverage these insights to implement targeted interventions, such as enhanced Maritime English/SMCP proficiency programs (with feedback/closed-loop drills), fatigue/stress management, and watch scheduling to reduce cognitive load, and protocols tailored for high-pressure, high-complexity operations (e.g. standardized phrasing/check-back under time pressure). Organizational levers, recurrent training, safety culture reinforcement, and language-aware watch composition/role allocation can further reduce communication role violations, particularly on multilingual bridges and during adverse operating windows.

Despite its contributions, this study has several limitations that should be considered when interpreting the findings. First, the dataset is drawn from formal investigation reports issued by a finite set of authorities between 2012 and 2022, which are more likely to document higher-severity events; non-severe incidents and near misses are probably under-reported, potentially biasing the estimated relationships toward more serious cases. Second, although the narratives were processed using an NLP/LLM pipeline and expert review, the source reports themselves may contain reporting and linguistic biases, for example, variation in how investigators describe language problems, or how non-native English interactions are summarized in English-language texts, which can affect the consistency with which constructs such as proficiency, cognitive load, or safety culture are captured. Third, the model is calibrated only on accidents where language and communication issues were explicitly identified as causal or contributory, so the findings should not be generalized to all maritime accidents or to scenarios dominated by purely technical failures. Finally, because the SEM is estimated on historical reports under the regulatory, technological, and crewing conditions of 2012–2022, its generalizability to other regions, flag states not covered by the underlying databases, or future contexts with different communication technologies and organizational practices should be treated with caution. Future research could mitigate these limitations by incorporating additional reporting systems, systematically sampling less severe events, and combining narrative-derived indicators with real-time operational or monitoring data.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program [Grant Agreement No. 864724].

ORCID

Rafi Ullah Khan http://orcid.org/0000-0002-6887-3487

References

Ana, B.-V. 2011. "The Language of Seafaring: Standardized Conventions and Discursive Features in Speech Communications." *International Journal of English Studies* 11 (1): 35–53. https://doi.org/10.6018/ijes/2011/1/137091.

Ahmmed, R. 2020. "The Discrepancy Between Standardised Communication Patterns and the Real-Life Conversations of Vessel Traffic Service: A Case Study in Chittagong Port, Bangladesh." *WMU Journal of Maritime Affairs* 19 (4): 509–532. https://doi.org/10.1007/s13437-020-00219-7.

Albert, E. C. 2024. "Analysing the Impact of Human Factors on Marine Accidents." *International Journal of Research and Innovation in Applied Science* 9 (11): 272–279.

Antão, P., and C. Guedes Soares. 2019. "Analysis of the Influence of Human Errors on the Occurrence of Coastal Ship Accidents in Different Wave Conditions Using Bayesian Belief Networks." *Accident Analysis and Prevention* 133:105262. https://doi.org/10.1016/j.aap.2019.105262.

Arenius, M., G. Athanassiou, and O. Sträter. 2010. "Systemic Assessment of the Effect of Mental Stress and Strain on Performance in a Maritime Ship-Handling Simulator." *IFAC Proceedings Volumes* 43 (13): 43–46. https://doi.org/10.3182/20100831-4-FR-2021.00009.

Argüelles, R. P., M. Bartolomé, J. A. García Maza, and F. Mateos Martín. 2021. "Ship-to-Ship Dialogues and Agreements for Collision Risk Reduction." *The Journal of Navigation* 74 (5): 1039–1056. https://doi.org/10.1017/S0373463321000448.

Austin, P. C., and E. W. Steyerberg. 2015. "The Number of Subjects per Variable Required in Linear Regression Analyses." *Journal of Clinical Epidemiology* 68 (6): 627–636. https://doi.org/10.1016/j.jclinepi.2014.12.014.

Bagozzi, R. P., and Y. Yi. 1989. "On the Use of Structural Equation Models in Experimental Designs." *Journal of Marketing Research* 26 (3): 271–284. https://doi.org/10.1177/002224378902600302.

Baygi, F., A. Smith, N. Mohammadian Khonsari, F. Mohammadi-Nasrabadi, Z. Mahmoodi, A. Mahdavi-Gorabi, and M. Qorbani. 2022. "Seafarers' Mental Health Status and Life Satisfaction: Structural Equation Model." *Frontiers in Public Health* 10:- 2022. https://doi.org/10.3389/fpubh.2022.969231.

Bialyk, O. Y. 2017a. "IMO Standard Marine Communication Phrases as a Part of Maritime Education and Training." *Modern Philology: Relevant Issues and Prospects of Research* 10/2017:76.

Bialyk, O. Y. 2017b. "The Role of IMO Standard Marine Communication Phrases in Safety of Navigation at Sea." *Науковий вісник Херсонського державного університету. Серія «Германістика та міжкультурна комунікація»* 3 (3/2017): 116–119.

Bošnjak, R., M. Bukljaš, D. Medić, and S. Vukša. 2021. "Proposal of New Measures for Risk Assessment in Navigation–A Case Study of the M/V Cosco Busan Accident." *Zeszyty Naukowe Akademii Morskiej w Szczecinie* 67 (139): 9–15.

Boström, M. 2020a. "Creating Clarity and Managing Complexity Through Co-Operation and Communication: The Case of Swedish Icebreaker Operations."

Boström, M. 2020b. "Mind the Gap! A Quantitative Comparison Between Ship-to-Ship Communication and Intended Communication Protocol." *Safety Science* 123:104567. https://doi.org/10.1016/j.ssci.2019.104567.

Boström, M. 2021. "Other-Initiated Repair as an Indicator of Critical Communication in Ship-to-Ship Interaction." *Journal of Pragmatics* 174:78–92. https://doi.org/10.1016/j.pragma.2021.01.007.

Byrne, B. M. 2013. *Structural Equation Modeling with Mplus: Basic Concepts, Applications, and Programming*. routledge.

Celik, M., and S. Cebi. 2009. "Analytical HFACS for Investigating Human Errors in Shipping Accidents." *Accident Analysis and Prevention* 41 (1): 66–75. https://doi.org/10.1016/j.aap.2008.09.004.

Chauvin, C. 2011. "Human Factors and Maritime Safety." *The Journal of Navigation* 64 (4): 625–632. https://doi.org/10.1017/S0373463311000142.

Chauvin, C., S. Lardjane, G. Morel, J.-P. Clostermann, and B. Langard. 2013. "Human and Organisational Factors in Maritime Accidents: Analysis of Collisions at Sea Using the HFACS." *Accident Analysis and Prevention* 59:26–37. https://doi.org/10.1016/j.aap.2013.05.006.

Chen, S.-Y., C.-S. Lu, K.-D. Ye, K.-C. Shang, J.-L. Guo, and J.-M. Pan. 2023. "Enablers of Safety Citizenship Behaviors of Seafarers: Leader-Member Exchange, Team-Member Exchange, and Safety Climate." *Maritime Policy & Management* 50 (1): 81–96. https://doi.org/10.1080/03088839.2021.1959077.

Committee, M. S. 2023. *Proposal for a New Output for Development of an MSC Circular to Address Time Pressure and Related Organizational Factors*. International Maritime Organization.

Comrey, A. L., and H. B. Lee. 2013. *A First Course in Factor Analysis*. Psychology press.

Crestelo Moreno, F., J. Roca Gonzalez, J. Suardíaz Muro, and J. A. García Maza. 2022. "Relationship Between Human Factors and a Safe Performance of Vessel Traffic Service

Operators: A Systematic Qualitative-Based Review in Maritime Safety." *Safety Science* 155:105892. https://doi.org/10.1016/j.ssci.2022.105892.

Gaweesh Sherif, M., U. Ahmed Irfan, and M. Ahmed Mohamed. 2024. "Assessment of Large Trucks Crash Severity on a Rural Interstate Road in Wyoming Using Decision Trees and Structural Equation Model." *Journal of Transportation Engineering, Part A: Systems* 150 (1): 05023008. https://doi.org/10.1061/JTEPBS.TEENG-7446.

Goerlandt, F., and H. Liu. 2024. "Readability of Maritime Accident Reports: A Comparative Analysis." *Maritime Policy & Management* 51 (7): 1410–1422. https://doi.org/10.1080/03088839.2023.2166685.

Hair, J. F. 2009. *Multivariate Data Analysis*. 7th ed. Pearson.

Hair, J. F., J. J. Risher, M. Sarstedt, and C. M. Ringle. 2019. "When to Use and How to Report the Results of PLS-SEM." *European Business Review* 31 (1): 2–24. https://doi.org/10.1108/EBR-11-2018-0203.

Haryani, M., H. R. Januarius, and Y. Issy. 2023. "The Essential of Using Standard Marine Communication Phrases in Avoiding Accidents at Sea for Indonesian Seafarers." *International Journal of Education, Vocational and Social Science* 2 (2): 51–67. https://doi.org/10.99075/ijevss.v2i02.216.

Hasanspahić, N., S. Vujičić, V. Frančić, and L. Čampara. 2021. "The Role of the Human Factor in Marine Accidents." *Journal of Marine Science and Engineering* 9 (3): 261. https://doi.org/10.3390/jmse9030261.

Hetherington, C., R. Flin, and K. Mearns. 2006. "Safety in Shipping: The Human Element." *Journal of Safety Research* 37 (4): 401–411. https://doi.org/10.1016/j.jsr.2006.04.007.

Horck, J. 2008. "Cultural and Gender Diversities Affecting the Ship/Port Interface." First International Ship Port Interface Conference (ISPIC 2008), Bremen, Germany. Retrieved from May. http://bit.ly/2gowXYY.

Hu, S., Z. Li, Y. Xi, X. Gu, and X. Zhang. 2019. "Path Analysis of Causal Factors Influencing Marine Traffic Accident via Structural Equation Numerical Modeling." *Journal of Marine Science and Engineering* 7 (4): 96. https://doi.org/10.3390/jmse7040096.

John, P., B. Brooks, and U. Schriever. 2017. "Profiling Maritime Communication by Non-Native Speakers: A Quantitative Comparison Between the Baseline and Standard Marine Communication Phraseology." *English for Specific Purposes* 47:1–14. https://doi.org/10.1016/j.esp.2017.03.002.

Khan, R. U., J. Yin, F. Shair Mustafa, and A. Omar Ahmad Farea. 2021. "A Data Centered Human Factor Analysis Approach for Hazardous Cargo Accidents in a Port Environment." *Journal of Loss Prevention in the Process Industries* 75:104711. https://doi.org/10.1016/j.jlp.2021.104711.

Khan, R. U., J. Yin, F. Shair Mustafa, and S. Wang. 2022. "Analyzing Human Factor Involvement in Sustainable Hazardous Cargo Port Operations." *Ocean Engineering* 250:111028. https://doi.org/10.1016/j.oceaneng.2022.111028.

Knapp, S., and M. van de Velden. 2024. "Exploration of Machine Learning Methods for Maritime Risk Predictions." *Maritime Policy & Management* 51 (7): 1443–1473. https://doi.org/10.1080/03088839.2023.2209788.

Koenig, K. L., and C. H. Schultz. 2016. *Koenig and Schultz's Disaster Medicine: Comprehensive Principles and Practices*, *Cambridge medicine*: Cambridge University Press.

Lan, H., S. Wang, and W. Zhang. 2024. "Predicting Types of Human-Related Maritime Accidents with Explanations Using Selective Ensemble Learning and SHAP Method." *Heliyon* 10 (9): e30046. https://doi.org/10.1016/j.heliyon.2024.e30046.

Lee, J.-Y., J.-H. Chung, and B. Son. 2008. "Analysis of Traffic Accident Size for Korean Highway Using Structural Equation Models." *Accident Analysis and Prevention* 40 (6): 1955–1963. https://doi.org/10.1016/j.aap.2008.08.006.

Li, G., K. Gao, J. Weng, and X. Qu. 2025. "Deciphering Spatial Heterogeneity of Maritime Accidents Considering Impact Scale Variations." *Maritime Policy & Management* 52 (2): 309–333. https://doi.org/10.1080/03088839.2024.2347893.

Li, Y. 2019. "An Empirical Research on Chinese Seafarers' Communicative Competence Training and Assessment Methods." In *Masters in Science, Maritime Safety Environmental Management*, World Maritime University.

Liu, D., and L. Cheng. 2024. "MAKG: A Maritime Accident Knowledge Graph for Intelligent Accident Analysis and Management." *Ocean Engineering* 312:119280. https://doi.org/10.1016/j.oceaneng.2024.119280.

Liu, L., R. U. Khan, M. Afzaal, and M. Asad. 2026. "A Corpus-Based Quantitative Risk Assessment of Language Barriers in Maritime Safety." *Maritime Policy & Management* 53 (3): 524–558. https://doi.org/10.1080/03088839.2025.2507218.

Lomax, R. G. 2004. *A Beginner's Guide to Structural Equation Modeling*. psychology press.

Lützhöft, M., M. R. Grech, and T. Porathe. 2011. "Information Environment, Fatigue, and Culture in the Maritime Domain." *Reviews of Human Factors and Ergonomics* 7 (1): 280–322. https://doi.org/10.1177/1557234X11410391.

Lützhöft, M. H., and S. W. A. Dekker. 2002. "On Your Watch: Automation on the Bridge." *The Journal of Navigation* 55 (1): 83–96. https://doi.org/10.1017/S0373463301001588.

Ma, L., X. Ma, H. Lan, Y. Liu, and W. Deng. 2022. "A Data-Driven Method for Modeling Human Factors in Maritime Accidents by Integrating DEMATEL and FCM Based on HFACS: A Case of Ship Collisions." *Ocean Engineering* 266:112699. https://doi.org/10.1016/j.oceaneng.2022.112699.

Maglić, L., S. Valčić, A. Gundić, and L. Maglić. 2020. "Voice Communication Systems Impact on Navigating Officers." *Journal of Marine Science and Engineering* 8 (3): 197. https://doi.org/10.3390/jmse8030197.

Mallam, S. C., S. Nazir, and A. Sharma. 2020. "The Human Element in Future Maritime Operations – Perceived Impact of Autonomous Shipping." *Ergonomics* 63 (3): 334–345. https://doi.org/10.1080/00140139.2019.1659995.

Marsh, H. W., and D. Hocevar. 1988. "A New, More Powerful Approach to Multitrait-Multimethod Analyses: Application of Second-Order Confirmatory Factor Analysis." *Journal of Applied Psychology* 73 (1): 107. https://doi.org/10.1037/0021-9010.73.1.107.

Miklody, D., W. M. Uitterhoeve, D. van Heel, K. Klinkenberg, and B. Blankertz. 2016. "Maritime cognitive workload assessment." In *International Workshop on Symbiotic Interaction*, 102–114. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-57753-1.

Mönnigmann, B., and A. Čulić-Viskota. 2017. "Standardised English Language Proficiency Testing for Seafarers." *Transactions on Maritime Science* 6 (2): 147–154. https://doi.org/10.7225/toms.v06.n02.007.

Noble, A. 2015. "The IMO SMCP 15 Years on Current Perceptions and Realistic Recommendations." IMEC27 Proceedings, 127–145.

Ofori-Boateng, R., M. Aceves-Martins, N. Wiratunga, and C. Francisco Moreno-Garcia. 2024. "Towards the Automation of Systematic Reviews Using Natural Language Processing, Machine Learning, and Deep Learning: A Comprehensive Review." *Artificial Intelligence Review* 57 (8): 200. https://doi.org/10.1007/s10462-024-10844-w.

Othman, M. R., and E. Adolf Naintin. 2016. "The Relationship Between Maritime Education and Employer Trust: The Structural Equation Modelling (SEM) Perspective." *WMU Journal of Maritime Affairs* 15 (2): 293–316. https://doi.org/10.1007/s13437-015-0093-1.

Pejaković, S. K. 2014. "Pragmatic Aspect of Maritime English Language." 2014. *Journal of Educational and Social Research*. https://doi.org/10.5901/jesr.2014.v4n4p512.

Ricketts, J., D. Barry, W. Guo, and J. Pelham. 2023. "A Scoping Literature Review of Natural Language Processing Application to Safety Occurrence Reports." *Safety* 9 (2): 22. https://doi.org/10.3390/safety9020022.

Ronca, V., E. Uflaz, O. Turan, H. Bantan, S. N. MacKinnon, A. Lommi, S. Pozzi, et al. 2023. "Neurophysiological Assessment of an Innovative Maritime Safety System in Terms of Ship Operators' Mental Workload, Stress, and Attention in the Full Mission Bridge Simulator." *Neurophysiological Assessment of an Innovative Maritime Safety System in Terms of Ship*

Operators' Mental Workload, Stress, and Attention in the Full Mission Bridge Simulator. Brain Sciences 13 (9): 1319. https://doi.org/10.3390/brainsci13091319.

Safety4Sea. 2020. "Language Differences Onboard: A Barrier to Ship Safety?" Safety4Sea, Accessed 10th. March. https://safety4sea.com/cm-language-differences-onboard-a-barrier-to-ship-safety/#:~:text=Did%20you%20know%3F .

Satler, T. B., and V. Jurkovič. 2024. "Solving Maritime Communication Challenges with Digimar: A Practical Approach." *Maritime Transport Conference* (10).

Seo, D.-C. 2005. "An Explicative Model of Unsafe Work Behavior." *Safety Science* 43 (3): 187–211. https://doi.org/10.1016/j.ssci.2005.05.001.

Seo, H.-C., Y.-S. Lee, J.-J. Kim, and N.-Y. Jee. 2015. "Analyzing Safety Behaviors of Temporary Construction Workers Using Structural Equation Modeling." *Safety Science* 77:160–168. https://doi.org/10.1016/j.ssci.2015.03.010.

Setiawan, A., and A. Razak Abdul Hadi. 2025. "Analysis of Human Element Factors Based on Big Data Impact of the Work Environment on Employee Stress and Fatigue Levels in the Maritime Industry." *Analysis of Human Element Factors Based on Big Data Impact of the Work Environment on Employee Stress and Fatigue Levels in the Maritime Industry (February 03, 2025)*. https://doi.org/10.2139/ssrn.5150085.

Sheykhfard, A., and F. Haghighi. 2020. "Driver Distraction by Digital Billboards? Structural Equation Modeling Based on Naturalistic Driving Study Data: A Case Study of Iran." *Journal of Safety Research* 72:1–8. https://doi.org/10.1016/j.jsr.2019.11.002.

Sheykhfard, A., F. Haghighi, T. Nordfjærn, and M. Soltaninejad. 2020. "Structural Equation Modelling of Potential Risk Factors for Pedestrian Accidents in Rural and Urban Roads." *International Journal of Injury Control and Safety Promotion* 28 (1): 46–57. https://doi.org/10.1080/17457300.2020.1835991.

Sheykhfard, A., S. Jones, E. Sadeghvaziri, S. Koppel, S. Das, and M. Nankali. 2025. "Integrating Empirical and Subjective Evidence on Young Drivers' Risk Perceptions and Crash Factors." *Transportation Research Record* 2679 (11): 892–908. https://doi.org/10.1177/03611981251350648.

Shi, X., D. Xu, H. Zhuang, and C. Liu. 2022. "How Unsafe Acts Occur: An Automatic Text Mining Study." *Maritime Policy & Management* 49 (6): 820–830. https://doi.org/10.1080/03088839.2021.1894363.

Simões-Marques, M., A. Frias, and P. B. Água. 2021. "Human Factors Impact in the Security and Safety of the Maritime Domain." In *Advances in Human Factors and System Interactions. AHFE 2021. Lecture Notes in Networks and Systems*, edited by I. L. Nunes, Vol. 265. Cham: Springer. https://doi.org/10.1007/978-3-030-79816-1_4.

Stevens, J. 2002. *Applied Multivariate Statistics for the Social Sciences*. Vol. 4. Lawrence erlbaum associates Mahwah, NJ.

Ullman, J. B., and P. M. Bentler. 2012. "Structural Equation Modeling." Handbook of Psychology, *Second Edition*.

Vagias, N. 2010. "A Bayesian Network Application for the Prediction of Human Fatigue in the Marine Industry." Unpublished dissertation, National Technical University of Athens.

Vivek, A. K., S. Gupta, T. Khan, and S. Sourava Mohapatra. 2024. "Strategies to Mitigate Safety and Associated Problems at Gated Rail Road Grade Crossing: A Structural Equation Modelling Approach." *Transport Policy* 146:19–30. https://doi.org/10.1016/j.tranpol.2023.11.002.

Wahid, A., M. Y. Jinca, T. Rachman, and J. Malisan. 2024. "Influencing Factors of Safety Management System Implementation on Traditional Shipping." *Sustainability* 16 (3): 1152. https://doi.org/10.3390/su16031152.

Wang, H., N. Chen, C. G. S. Jinhui Jiang, B. Wu, and B. Wu. 2025. "A Data-Driven AcciMap-BN Model for Risk Factor Analysis of Collisions Between Merchant Ships and Fishing Vessels." *Maritime Policy & Management*: 1–31. https://doi.org/10.1080/03088839.2025.2524533.

Wang, Y., and S. Fu. 2022. "Framework for Process Analysis of Maritime Accidents Caused by the Unsafe Acts of Seafarers: A Case Study of Ship Collision." *Journal of Marine Science and Engineering* 10 (11): 1793. https://doi.org/10.3390/jmse10111793.

Xu, T., Y. Xiao, and Z. Jiang. 2021. "Maritime Pilots' Risky Operational Behavior Analysis Based on Structural Equation Model." *Discrete Dynamics in Nature & Society* 2021 (1:3611859. https://doi.org/10.1155/2021/3611859.

Yin, J., R. Ullah Khan, M. Afzaal, H. M. Almalki, M. Ahmad Saleem Khasawneh, and S. Al Sulaie. 2025. "Quantitative Risk Assessment of Speech Acts and Lexical Factors in Maritime Communication Failures and Accidents." *Safety Science* 191:106968. https://doi.org/10.1016/j.ssci.2025.106968.

Yin, J., R. Ullah Khan, R. N. Muhammad Afzaal, X. Shanshan, A. Jamal, and A. Jamal. 2024. "A Fuzzy Bayesian Quantitative Risk Assessment for Language and Communication Induced Accidents in Maritime Operations." *Ocean & Coastal Management* 259:107449. https://doi.org/10.1016/j.ocecoaman.2024.107449.

Yu, J., J. Huang, G. Liu, and G. Zhang. 2025. "Improving Seafarers' Job Satisfaction and Retention Intention: A Job Demands-Resources Perspective." *Maritime Policy & Management*: 1–31. https://doi.org/10.1080/03088839.2025.2532462.

Žagar, D., M. Svetina, A. Košir, and F. Dimc. 2020. "Human Factor in Navigation: Overview of Cognitive Load Measurement During Simulated Navigational Tasks." *Journal of Marine Science and Engineering* 8 (10): 775. https://doi.org/10.3390/jmse8100775.

Zhou, Y., Z. Haque Munim, M. Li, J. Zhang, Y. Li, and X. Feng. 2025. "A Systematic Review on Human-Centred Maritime Operations Analytics Involving Psychophysiological Data." *Maritime Policy & Management*: 1–28. https://doi.org/10.1080/03088839.2025.2548793.

Ziegler, N. 2024. "Miscommunication in Maritime Contexts: Findings from Phase 1 and 2." *Prince William Sound Regional Citizens' Advisory Council*.