WMU J Marit Affairs (2018) 17:71–90
https://doi.org/10.1007/s13437-018-0136-5

ARTICLE

Consistency in the development of performance assessment methods in the maritime domain

Jørgen Ernstsen[1] · **Salman Nazir**[1]

Received: 22 December 2016 / Accepted: 22 January 2018 / Published online: 14 February 2018

Abstract The maritime industry is considered to be a backbone of the global economy. It is therefore imperative to ensure that maritime operations run safely and efficiently. Assessment of maritime performance is necessary for designers and engineers to be able to pinpoint the weakest links in the system and make impactful system improvements. The current article presents a systematic quantitative literature review of research on performance assessment in the maritime industry with the goal of establishing an understanding of accuracy and consistency in the development of methods used to assess performance. The review focuses on four major segments within the industry—port logistics, ship handling, safety and environmental research—and investigates their uses in developing accurate and consistent performance assessment methods. After the completion of an exclusion process, 62 articles published in a wide range of academic journals were used in the analysis. Two important conclusions were drawn from the analysis. First, performance assessment is generally consistent throughout the maritime industry; most papers used accurate and consistent approaches to develop the methods ($n = 43$). A subsequent bivariate analysis revealed a call for increased attention to the development of assessment methods within the maritime segment of ship handling. The current study suggests and discusses certain directions with regard to assessment research in the maritime industry.

Keywords Maritime performance · Assessment · Performance indicators · Literature review · Complex operations

✉ Jørgen Ernstsen
jorgen.ernstsen@usn.no

✉ Salman Nazir
salman.nazir@usn.no

[1] Training and Assessment Research Group, Department of Maritime Operations, University College of Southeast Norway, Borre, Norway

1 Introduction

The maritime industry is massive and is responsible for over 90% of global trade. The industry employs around 1,653,500 people across many countries (BIMCO 2015). This responsibility requires many high-stakes operations to ensure that goods are transported across the globe in a timely manner. High-stakes operations are complex and system deviations can have devastating consequences. The complexity of such operations is associated interdependent collaborations and dynamic decision-making; these factors can add up to make the work of maritime operators exceptionally straining (Kluge 2014). This complexity has led to severe accidents such as the capsizing of *Costa Concordia*, the Sewol ferry tragedy (Kim et al. 2016) and the El Faro accident (Coast Guard 2017).

The potential consequences following errors in high-stakes operations are costly in terms of environmental damages, operating expenses and health hazards (Naderpour et al. 2015). Accidents are believed to be an inevitable part of high-stakes operations; such as the *Federal Kivalina*, *Crete Cement* and *M/V Godafoss* accidents have repeatedly demonstrated this reality (Accident Investigation Board 2010a, b, 2012). In response, many significant incentives exist for ship owners, crews and local communities to identify which measures that can prevent accidents or mitigate damages. Several possibilities exist for advancing on this issue such as improving technical systems, designs, or the engineering phase. Other options include establishing training and hiring procedures (Leveson 2011).

Regardless of which system components are inspected, performance assessment remains an essential method for identifying the measures that strengthen safety and efficiency (Wiggins 1993). However, it is difficult to assess high-stakes operational performance (Delandshere and Petrosky 1998). Accurate and consistent performance assessments are necessary to provide useful information regarding safety and efficiency. This need implies a systematic effort to understand the mechanisms in an operation in order for an assessor to pinpoint specific parts of the system that require enhancements (Bouejla et al. 2014). Information about system weaknesses is crucial in order to apply improvements that will eventually lead to safer and more efficient operations. This systematic effort requires that the assessment tool is able to capture key nuances in operations. The benefits of powerful assessment tools are even larger when the consequences are greater.

Assessment methods can take many forms. Generally, they consist of a hierarchy of previously identified performance indicators in which higher indicators are calculated based on lower indicators (Ernstsen et al. 2016); Manca et al. (2014) provide an example. The quality of an assessment depends on an accurate and consistent development of the method and proper identification of performance indicators. Consistency in this development process tends to vary in other industries and operational segments (Aditya et al. 2015).

Considering the maritime industry's indispensable position in global trade, comprising a plethora of high-stakes and challenging operations, it is necessary to appraise the integrity of the way performance is assessed. Accurate and consistent performance assessment benefits all parties and leads to higher returns in terms of operational safety and efficiency, as demonstrated in other high-stakes domains, such as aviation (Mavin and Roth 2014), railways (Abril et al. 2008) and power plants (Nazir et al. 2014; Nazir et al. 2015).

The lack of performance assessment research on maritime operations is alarming, but attention to the matter has increased in recent decades (Rødseth et al. 2016). To maintain momentum in producing new research, the current study investigates the accuracy and consistency of performance assessment methods used in the maritime industry by examining how these methods are developed.

The accuracy and consistency of developing performance assessments across four major maritime segments were investigated using a systematic quantitative literature review to identify performance methods and the approaches used in all examined research papers. The following section presents a theoretical overview of the four approaches for developing performance assessment methods and a description of the maritime segments that are investigated in the current research. Subsequently is a presentation of the method; it is followed by a presentation of the results and analysis. The paper concludes with a discussion of the results.

2 Four approaches for developing performance assessment methods

Evaluating the development process of a tool provides information about its accuracy and consistency (Downing 2003). Following this argument, bottom-up, top-down and hybrid approaches can be considered to be accurate and consistent, while disconnections between data or theory and application are generally associated with inaccurate and inconsistent approaches to developing performance assessment methods (Hinkin et al. 1997). See Fig. 1 below.

2.1 Bottom-up approach

Studies that fit in this category have a goal of developing or identifying performance indicators (PIs) within a defined operation or industry. The methods of finding PIs vary, though most involve using interviews, questionnaires and observations from subject matter experts (SMEs). The work commonly generates a list of PIs that are specific to

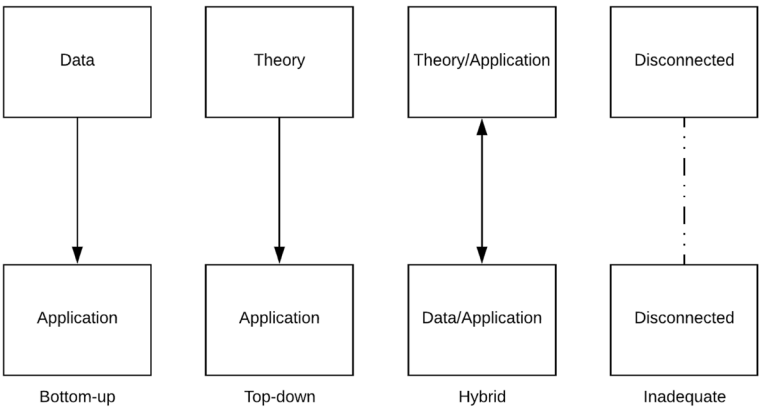

Fig. 1 Four approaches to performance assessment. Bottom-up, top-down and hybrid approaches are considered to be accurate and consistent processes for developing performance assessment methods

the operation but can also be developed for generic usage; Leriche et al. (2015) provide an example.

Considering that the PIs identified are not limited by existing methods can be advantageous since the researchers have flexibility to adapt to the assessment for a specific situation. On the other hand, disadvantages raise questions of validity, such as whether data collection and subsequent analyses have been properly designed and carried out. It can be challenging to develop PIs in sociotechnical systems without the assistance of a theoretical framework. For example, some PIs may count a performance score twice at different stages of an operation. It is possible for a theoretical framework to account for this misinterpretation using algebraic calculations or through sophisticated modelling prior to measurement, though the framework's calculations must be valid as well.

2.2 Top-down approach

Another approach is to use established literature, theories, regulations, legislations and frameworks to assess the PIs associated with an operation. Studies with a top-down focus use an established PI framework to evaluate the performance of an operation. In addition, such research can provide further validation of PIs that were previously identified in studies that use the bottom-up approach, such as Talley et al. (2014).

Efficiency and validity are advantages of a top-down approach. In many situations, established frameworks can provide valuable definitions and formulas to effectively measure the performance of an operation, eliminating the work of developing new PIs. Furthermore, robust legislation, regulation and standardisation of measurement systems may justify a framework's validity and increase the trust of true measurement. One disadvantage is the lack of flexibility; if an established framework is tailored to a specific operation, the framework may condition the validity in another operation. Attention to and knowledge of a framework is necessary to use it effectively across situations and ensure a truer measurement of performance.

2.3 Hybrid approach

A combination of a bottom-up and a top-down approach can also be used. In this approach, data is gathered and analysed to develop PIs; at the same time, the PIs are evaluated against a set of predefined performance assessment frameworks; Sleire and Dale (2009) provide an example. The approach demands more resources than a single bottom-up or top-down approach, but it benefits from flexibility and established validity.

2.4 Inadequate approaches

Another approach is to haphazardly (or at least highly subjectively) determine a set of indicators for measuring performance. Depending on the available measurement tools, resources and knowledge of the system, this approach questions the validity of the data. One reason is that only a fraction of the system is measured, and interconnections existing in complex high-stakes operations are disregarded. Clearly, measuring all variables in a complex system is ideal, but a systematic approach may reveal the most important aspects of system performance. On the other hand, efficiently selecting

indicators makes it possible to pinpoint areas of focus and relevant variables in a system; however, the highly subjective selection of indicators may compromise the accuracy and consistency of the overall operation. The current paper refers to these approaches to developing assessment methods as *inadequate approaches*.

Comparing various development processes across research studies is feasible and valuable in contrast to merely comparing specific assessment methods developed for distinct purposes. The relation between the processes is illustrated in Fig. 2 below, in which the bottom row (shaded area) represents the aims of the current research.

2.5 Four major maritime segments

The current research scrutinises the process of developing performance assessment tools in four major maritime segments. Port logistics, ship handling, safety and environmental performance are investigated because they all play a significant part in most shipping operations, and are thus widely researched. This information is analysed to deduce the accuracy and consistency of assessment methods in each of the respective segments.

Ports are essential hubs in maritime trade. Ports have become increasingly complex, evolving from a rudimentary place where cargo is handled to a functional element in the logistics chain that involves the flow of commodities, people and information (Roh et al. 2007). Extensive research has been conducted to develop assessment methods that capture the complex interplay among all agents in a port in order to find the best solution to port logistics.

Ship handling is the manoeuvring of a vessel, which encompass both technical seamanship skills and teamwork skills among crewmembers. Maritime operators must withstand a harsh and dynamic environment, often in isolation. This work is challenging, and measures must be taken to ensure that a crew has the skills required to accomplish necessary tasks.

Similarly, safety concerns are highly important in high-stakes industries and have been widely researched. The costs of a safety breach can be tremendous, so utmost care must be taken to increase safety. However, measuring safety is difficult because of its complexity, and many resources are invested in developing assessment frameworks for safety concerns.

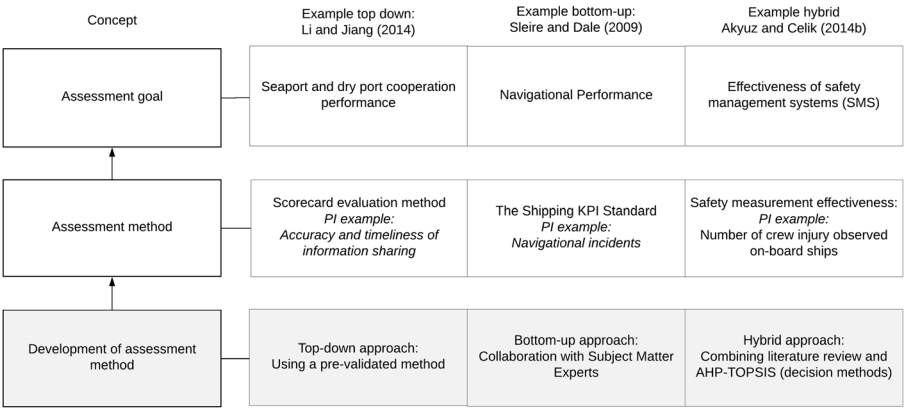

Fig. 2 How the aims of the current research (shaded area) fit into the overall process of assessing performance

Environmental considerations are increasingly relevant. Ship owners, local societies and governments are all apprehensive about the environment and express interest in green fleets. To be considerate of environmental impacts can yield productivity benefits for ship owners in the form of reduced fuel consumption as well as local benefits such as less pollution. Assessment methods have been developed to understand various aspects of the investigating environmental impact of operations from effects on coral reefs to carbon emissions.

3 Method

Peer-reviewed papers about maritime performance assessment were gathered from the Scopus, ScienceDirect and JSTOR databases. The criteria for including literature in the review follow the exclusion process depicted in Fig. 3 below. The time range considered was from 2005 to 2016; no relevant papers published before 2005 were identified.

3.1 Research statement and database search

A search statement was developed to ensure consistency across all database searches. The *use and development of performance indicators in the maritime industry* was broken down in four concepts; various combinations of these keywords (please see Table 1) have been explored in the literature. Concepts were topics in the search statement with relevant synonyms or alternative spellings such as maritime and marine, which are the British and American terms for the same concept that are both widely used in maritime literature. "Maritime" is the term used in the current paper. In total, 91 papers were found in Scopus, 568 were found in ScienceDirect (although search results only display 489 findings) and 44 were found in JSTOR. The same Boolean key-strain was used in all databases. After duplicates and unavailable papers (193 papers) were removed, 537 distinctive papers remained.

3.2 Process of exclusion

The subsequent step in the process involved excluding irrelevant papers. The papers were first excluded based on an evaluation of abstracts conducted according to the process depicted in Fig. 3. In the first exclusion process, 128 research papers were

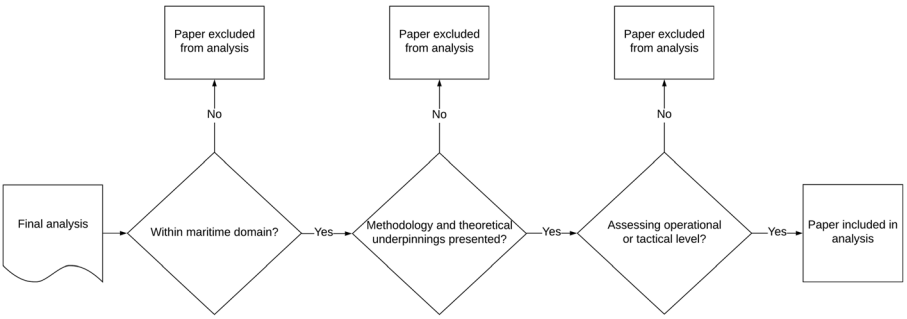

Fig. 3 Process of excluding papers in the literature review

Table 1 Keywords for the four concepts used to search for relevant literature

Concept 1	Concept 2	Concept 3	Concept 4
Performance indicators	Maritime	Framework	Method
Key performance indicators	Marine	Measure	Methodology
	(AND) shipping	reference model	

Boolean key-strain:
("performance indicators" OR "key performance indicators") AND (maritime OR marine) AND (framework OR measure OR reference model) AND (method OR methodology) AND shipping

selected for further examination. In the second part, complete articles were read to further assess relevance in relation to the search criteria; the same process was followed. Sixty-two research papers qualified from the second exclusion process and were chosen to be part of the literature evaluation. Complete numbers for each stage of the process are presented in Table 2 below.

The papers were evaluated based on their relevance to the maritime industry, whether methodology and theoretical underpinnings were presented in the paper and whether the performance assessments were at the operational or tactical level. Operational performance assessments evaluate how a vessel performs within an operation, as with docking, navigating or dynamic positioning; tactical operational assessments evaluate how well a vessel performs across operations. Strategic evaluations, which are excluded from the current paper, are concerned with how an entire fleet performs over time and involve several economic calculations that were considered to be too indirectly related to job performance to be included in the current study.

3.3 Structuring the literature

The findings in the literature review were structured according to the maritime segments; the research papers were coded from A to D. Port logistics (A) encompass logistics and vessel handling when approaching a port. Ship handling (B) measures operational performance on board a vessel including both technical and navigational efficiency. Safety (C) concerns performance frameworks that assess both antecedents and consequences of crises. Environmental performance (D) focuses on research measuring green performance and the development of green performance indicators. Every paper was assessed in relation to the assessment methods and specifically in terms of which approach (i.e. bottom-up, top-down or hybrid) was taken in the conducted research.

Table 2 Process of excluding research papers

Databases	Research papers before exclusion	Research papers after exclusion of abstract	Research papers after exclusion of entire paper
All (unique papers)	537 papers	128 papers	62 papers

3.4 Analysing the literature

All papers were included in univariate and bivariate analyses. The univariate analysis investigated the descriptive statistics concerning assessment methods used in the literature review. Another descriptive analysis was conducted on the distribution of the development approaches. Subsequently, to evaluate the consistency and accuracy of assessment methods, a cross-tabulation analysis was conducted on the use of the various development approaches across the four maritime segments.

4 Results

The results from the literature is organised into Table 3. It provides a list of the various approaches used in the examined research papers. The coding shown in the tables corresponds with and is used to identify the specific papers. For instance, code A1 corresponds with the paper titles "When it comes to container port efficiency, are all developing regions equal?" and the table illustrates that the assessment method was developed using a top-down approach.

The "X" marks the approach used in the respective research papers.

4.1 Result from the univariate analysis

Two descriptive analyses were performed to determine the distribution of data. First, descriptive statistics for the assessment methods identified in the literature are presented. Seventeen unique performance assessment methods were identified, though some were adapted for specific settings. Eleven undefined and unique methods were catalogued in the review. Such methods are often associated with an inadequate development approach.

The second descriptive analysis focused on the frequency with which different approaches were used to develop assessment methods. The top-down approach was the most prevalent approach (mode = 21 (34%)); bottom-up was the least-applied approach to performance assessment (15%). Combining adequate approaches (those that were consistent and accurate) revealed that 69.4% of the papers reviewed based their assessments on adequate research approaches. This finding signals an overall strong consistency for the maritime industry (Fig. 4).

4.2 Result from the bivariate analysis

A bivariate cross-tabulation analysis was performed to further investigate accuracy and consistency. For port logistics and ship handling, the *inadequate* approach was most dominant at 34% for port logistics and 46% for ship handling. The *hybrid* approach was most prevalent in safety assessments (50%), and the *top-down* approach was used most often in assessments of environmental concerns (58%). Table 4 presents the cross-tabular bivariate analysis of the assessment approaches and maritime segments. The analysis revealed that the majority of papers in all segments used adequate approaches.

The distribution of various assessment approaches was then analysed to pinpoint the accuracy and consistency of the assessment methods for each maritime segment. Each

Table 3 Listing assessment approaches that are used in the respective papers

Code	Title	Reference	Assessment development approach			
			Inadequate	Bottom-up	Top-down	Hybrid
A1	When it comes to container port efficiency, are all developing regions equal?	Suárez-Alemán et al. (2016)	–	–	X	–
A2	Measuring lean port performance	Marlow and Paixão Casaca (2003)	–	–	–	X
A3	A logistics and supply chain management approach to port performance measurement	Bichou and Gray (2004)	–	–	–	X
A4	Proposal for a flexible discrete event simulation model for assessing the daily operation decisions in a Ro–Ro terminal	Iannone et al. (2016)	X	–	–	–
A5	Identification of Occupational Health, Safety, Security (OHSS) and environmental performance indicators in port areas	Antão et al. (2016)	–	–	–	X
A6	Container Port Performance Measurement and comparison leveraging ship GPS traces and maritime open data	Chen et al. (2016)	–	–	X	–
A7	Exploring the drivers of port efficiency in Latin-America and the Caribbean	Serebrisky et al. (2016)	X	–	–	–
A8	Costs and benefits of speeding up reporting formalities in maritime transport	Vaghi and Lucietti (2016)	X	–	–	–
A9	Container terminal operations simulator (CTOS)—simulating the impact of extreme weather events on port operation	Chhetri et al. (2016)	X	–	–	–
A10	The revealed competitiveness of major ports in the East Asian region: an additive market share analysis.	Kim (2015)	X	–	–	–
A11	Simulating new logistics system of Le Havre Port	Leriche et al. (2015)	–	X	–	–
A12	Multi-objective genetic algorithm for berth allocation problem considering daytime preference	Hu (2015)	–	–	X	–
A13	Choosing optimal bunkering ports for liner shipping companies: a hybrid Fuzzy-Delphi-TOPSIS approach	Wang et al. (2014)	–	X	–	–
A14	Characteristics of European inland ports: a statistical analysis of inland waterway port development in Dutch municipalities	Wiegmans et al. (2015)	X	–	–	–

Table 3 (continued)

Code	Title	Reference	Assessment development approach			
			Inadequate	Bottom-up	Top-down	Hybrid
A15	A new risk quantification approach in port facility security assessment	Yang et al. (2014)	–	X	–	–
A16	Port service chains and port performance evaluation	Talley et al. (2014)	–	–	X	–
A17	Port capacity evaluation formula for general cargo	Park et al. (2014)	–	–	–	X
A18	Cooperation performance evaluation between seaport and dry port: case of Qingdao port and Xi'an port	Li and Jiang (2014)	–	–	X	–
A19	Port management performance and contextual variables: which relationship*? Methodological and empirical issues	Bergantino et al. (2013)	X	–	–	–
A20	Logistics resources in seaport performance: multi-criteria analysis	Da Cruz et al. (2013)	–	–	X	
A21	Intermodal connectivity as a port performance indicator	De Langen and Sharypova (2013)	–	–	–	X
A22	Operational performance and physical capacity of Iberian seaport: a multi-criteria analysis	Cruz et al. (2012)	–	–	X	–
A23	Assessment model of the port effectiveness and efficiency (case study: Western Indonesia Region)	Sutomo and Soemardjito (2012)	–	–	X	–
A24	Port evolution and performance in changing logistics environments	Woo et al. (2011)	–	–	–	X
A25	Green management practices and firm performance: a case of container terminal operations	Lun (2011)	X	–	–	–
A26	An evaluation of free trade port zone in Taiwan	Chiu et al. (2011)	–	X	–	–
A27	Measuring productivity and efficiency of major ports of India	Monteiro (2010)	X	–	–	–
A28	Waiting profiles: an efficient protocol for enabling distributed planning of container barge rotations along terminals in the port of Rotterdam.	Douma et al. (2009)	X	–	–	–
A29	Port choice and freight forwarders	Tongzon (2009)	–	–	–	X
B1	The Shipping KPI Standard.	Sleire and Dale (2009)	–	X	–	–

Table 3 (continued)

Code	Title	Reference	Assessment development approach			
			Inadequate	Bottom-up	Top-down	Hybrid
B2	Selecting technologies towards compliance with MARPOL Annex VI: the perspective of operators	Schinas and Stefanakos (2014)	–	–	X	–
B3	Benchmarking of CO2 transport technologies: part II—offshore pipeline and shipping to an offshore site	Roussanaly et al. (2014)	X	–	–	–
B4	A decision support system for mission-based ship routing considering multiple performance criteria	Dong et al. (2016)	X	–	–	–
B5	Multi-objective optimization for planning liner shipping service with uncertain port times	Song et al. (2015)	X	–	–	–
B6	Impacts of implementation of the effective maritime security management model on organizational performance of shipping companies	Sadovaya and Thai (2015)	–	–	X	–
B7	The impact of port operations on efficient ship operations from both economic and environmental perspectives	Moon and Woo (2014)	X	–	–	–
B8	Cruise carrying capacity: a conceptual approach	Stefanidaki and Lekakou (2014)	–	–	X	–
B9	Multi-criteria analysis of two CO2 transport technologies	Roussanaly et al. (2013a, b)	X	–	–	–
B10	Benchmarking of CO2 transport technologies: part I—onshore pipeline and shipping between two onshore areas.	Roussanaly et al. (2013a, b)	X	–	–	–
B11	Who benefits from simulator training: personality and heart rate variability in relation to situation awareness during navigational training	Saus et al. (2012)	–	X	–	–
B12	The northwest passage: a simulation	Somanathan et al. (2009)	–	–	X	–
B13	A multi-methodological approach for shipping registry selection in maritime transportation industry	Kandakoglu et al. (2009)	–	X	–	–
C1	Port safety and the container revolution: a statistical study on human factor and occupational accidents over the long period	Fabiano et al. (2010)	X	–	–	–
C2	Port safety evaluation from a captain's perspective: the Korean experience	Pak et al. (2015)	–	–	–	X

Table 3 (continued)

Code	Title	Reference	Assessment development approach			
			Inadequate	Bottom-up	Top-down	Hybrid
C3	Utilisation of cognitive map in modelling human error in marine accident analysis and prevention	Akyuz and Celik (2014b)	–	–	–	X
C4	A hybrid decision-making approach to measure effectiveness of safety management system implementation on-board ships	Akyuz and Celik (2014a)	–	–	–	X
C5	Identifying crucial safety assessment criteria for passenger ferry services	Lu and Tseng (2012)	–	–	X	–
C6	Evaluation of safety and environmental risk at individual ship and company level	Heij and Knapp (2012)	X	–	–	–
C7	Accident investigation in the Norwegian petroleum industry—common features and future challenges	Okstad et al. (2012)	–	–	X	–
C8	Sailing on Friday: developing the link between safety culture and performance in safety-critical systems	Grabowski et al. (2010)	–	–	–	X
D1	Environmental cost and eco-efficiency from vessel emissions in Las Palmas Port	Tichavska and Tovar (2015a)	–	–	X	–
D2	New environmental performance baseline for inland ports: a benchmark for the European inland port sector	Seguí et al. (2016)	–	–	–	X
D3	Port-city exhaust emission model: an application to cruise and ferry operations in Las Palmas Port	Tichavska and Tovar (2015b)	–	–	X	–
D4	Using indicators to evaluate the Taiwanese distant-water fishery-policy performance	Lin et al. (2014)	–	X	–	–
D5	Enhancing management effectiveness of environmental protected areas, Thailand	Satumanatpan et al. (2014)	–	–	X	–
D6	Greening logistics centers: the evolution of industrial buying criteria towards green	Altuntaş and Tuna (2013)	–	–	X	–
D7	Do firms get what they want from ISO 14001 adoption: an Australian perspective	Prajogo et al. (2012)	–	–	X	–

Table 3 (continued)

Code	Title	Reference	Assessment development approach			
			Inadequate	Bottom-up	Top-down	Hybrid
D8	The hard choices of sustainability	Standal and Utne (2011)	X	–	–	–
D9	System evaluation of sustainability in the Norwegian cod-fisheries	Utne (2007b)	–	–	X	–
D10	Performance indicators in MPA management: using questionnaires to analyze stakeholder preferences	Himes (2007)	–	X	–	–
D11	Development of a system of indicators for sustainable port management	Peris-Mora et al. (2005)	–	–	–	X
D12	Acceptable sustainability in the fishing fleet	Utne (2007a)	–	–	X	–

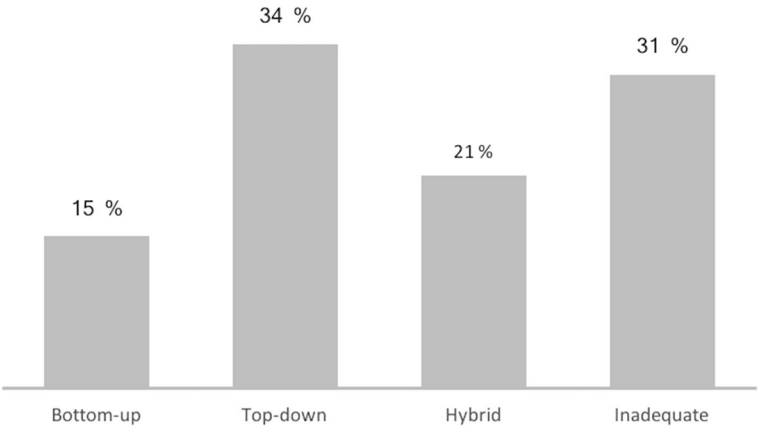

Fig. 4 Distribution of papers along the four research approaches

segment received a score based on the number of research papers addressing each of the approaches. Furthermore, each approach received a *weight* reflecting its impact on the development process; this weight was determined using four assessment research experts to ensure consistency. The bottom-up and top-down approaches were weighted at 1, the hybrid approach was weighted at 1.5 and the inadequate approach had a negative weight of − 0.5.

The weights were devised to favour the more extensive hybrid method and penalise the lack of an accurate and consistent approach. The result for each segment was a relative proportional score due to the uneven return of papers for each segment. The maximum score was achieved if all papers for a segment received a weight of 1.5, meaning that they used the hybrid method, and the relative score was the proportion of the score to the maximum score for each segment. The environmental segment received the highest score (0.64), port logistics and safety received the middle scores (0.40 and 0.58, respectively) and ship handling received the lowest score (0.21) (Table 5).

5 Discussion

A majority of the performance assessment research papers were found to develop assessment methods using adequate development approaches. At the same time, ship handling was suggested to receive increased attention with regards to consistency and accuracy in the development of assessment methods.

Table 4 Distribution of adequate and inadequate approaches with regards to each maritime segment

Approach	Port logistics	Ship handling	Safety	Environmental	Total:
Adequate	19	7	6	11	43
Inadequate	10	6	2	1	19

Table 5 Relative accuracy and consistency comparison with the corresponding approaches

Approach	Weight (w)	Port logistics	Ship handling	Safety	Environmental
Bottom-up	1.0	4	3	0	2
Top-down	1.0	8	4	2	7
Hybrid	1.5	10.5	0	6	3
Inadequate	− 0.5	− 5	− 3	− 1	− 0.5
Score		17.5	4	7	11.5
Maximum score		43.5	19.5	12	18
Relative score		0.40	0.21	0.58	0.64

The bivariate analysis found that environmental research returned the highest relative score of the four segments (0.64). Environmental research has received much attention in recent years (Chu et al. 2017), and newer research may have increased attention developing a comprehensive method for performance assessment. Another explanation may be that the maritime industry is suspected to have a high environmental footprint (Lam 2015), and strong environmental performance is a key interest for all stakeholders.

The port logistics and safety segments received the scores 0.40 (port logistics) and 0.58 (safety) for accuracy and consistency. Port logistics were associated with a relatively high number of papers using inadequate methods to develop assessment methods; however, the majority of papers used adequate approaches. Safety assessment research was associated with the highest percentage of papers applying the more extensive *hybrid* approach to developing assessment methods. This finding suggests that some attention should be shifted to using adequate tools to develop methods for assessing port logistics; momentum in safety research should be maintained.

Ship handling scored the lowest (0.21); this field concerns seamanship and social collaboration on vessels (Ernstsen et al. 2017). This maritime segment received the lowest score regarding accuracy and consistency in the development of assessment methods. Ship handling is difficult to measure and has mostly been measured by the examination of the technical parameters of vessel performance; Sleire and Dale (2009) provide an example. It may be perceived as less advantageous to use a comprehensive approach. Regardless, it can be argued that accurate and consistent measurements for ship handling are also beneficial (Bouejla et al. 2014). As the shift from manual to automated systems continues, it is imperative that assessments of vessel performance are accurate and consistent. The low score suggests that further research on the assessment of ship handling is necessary.

Assessing performance is critical to determining operational safety and efficiency in high-stakes operations. Evidence of inadequate performance assessments is apparent in the existing literature, as 31% of the papers examined used approaches classified as inaccurate and inconsistent. It is difficult and time-consuming to adequately develop performance methods, and it may even be a conscious and constructive decision for certain operations to adopt a pragmatic approach. Nevertheless, the current paper argues that inaccurate and inconsistent assessment methods may cause *more harm than good* if pragmatic approaches are portrayed and misperceived as absolute and reliable measures of a particular operation. It is essential to be conscious of the underlying

approaches used in the development of assessment methods for maritime operations; the current research emphasises this need.

It is worthy to mention some limitations. First, the Boolean logic applied impacted the research papers returned for analysis. The research papers in the review were examined carefully to ensure balance and proper representation of the literature. However, subjectivity was still present in the identification of relevant concepts and keywords used in the search string. This subjective effect was minimised by ongoing discussion among the researchers; however, it is still necessary to acknowledge this limitation. The identification of research concepts used as basis for the Boolean logic was also impacted by subjectivity, which influenced the subsequent identification of keywords and could have misled the study early on. Careful attention was paid to ensure that preconceived ideas and confirmation biases were minimised when the concepts for the systematic literature review were determined. Additionally, the exclusion criteria used to withdraw irrelevant research papers in systematic literature reviews influenced the results substantially. A step-by-step process was established to ensure that the literature was excluded in a consistent way, and the process was carefully verified in a dual review of the exclusion criteria. Finally, the use of a maximum score in the calculation can be considered misleading, as using a hybrid approach (which was required to achieve a maximum score) is not advisable or sensible in all circumstances and situations. However, the maximum score can be considered valuable for calculating the relative proportional score used to compare the respective maritime segments.

Although the findings of the current systematic literature review suggest an overall tendency to develop adequate assessment methods in the maritime industry, subsequent analyses of maritime segments and specified assessment approaches suggest opportunities for further improvement. For instance, it is suggested that standardising the way assessment methods are developed is further investigated. This could increase accuracy and consistency in the way performance is measured. It is also suggested that subsequent analyses pay increased attention to the development of ship handling performance frameworks. A comprehensive assessment framework to effectively determine ship-handling skills in high-stakes operations would make a significant contribution to maritime safety and efficiency.

6 Conclusion

The maritime industry is massive, and its vast impact on global ecology deserves to be accurately and consistently measured. The current study systematically investigated existing maritime literature to determine the prevalent use of consistent and accurate approaches to develop assessment methods. The findings suggest that assessment methods used in the maritime industry are developed using accurate and consistent approaches such as bottom-up, top-down and hybrid approaches. In the past, assessments of ship handling have commonly been using inadequate and highly subjective approaches to developing assessment methods. Therefore, it is proposed that the development of the methods used to assess performance in this maritime segment should receive additional attention. The current research paves the way for a systematic and increased understanding of performance assessment in the maritime industry. Currently, the authors are designing an experiment to evaluate the consistency and

accuracy of performance indicators for ship navigation with an aim to further increase the integrity of performance assessments and lead to a safer and more efficient industry.

References

Abril M, Barber F, Ingolotti L, Salido MA, Tormos P, Lova A (2008) An assessment of railway capacity. Transp Res E-Log 44(5):774–806. https://doi.org/10.1016/j.tre.2007.04.001

Accident Investigation Board, Norway (2010a) Crete Cement - IMO NO. 9037161, grounding at Aspond Island in the Oslo Fjord, Norway, on 19 November 2008. Report Sjø 1

Accident Investigation Board, Norway (2010b) Report on marine accident Federal Kivalina-IMO NO. 9205885 grounding at Årsundøya, Norway 6 October 2008. Report Sjø 1

Accident Investigation Board, Norway 2012 Report on investigation into marine accident M/V Godafoss V2PM7 grounding in Løperen, Hvaler on 17 February 2011. Report Sjø 1

Aditya P, Uday K, Diego G, Christer S (2015) Performance measurement and management for maintenance: a literature review. J Qual Maint Eng 21(1):2–33. https://doi.org/10.1108/JQME-10-2013-0067

Akyuz E, Celik M (2014a) A hybrid decision-making approach to measure effectiveness of safety management system implementations on-board ships. Saf Sci 68:169–179. https://doi.org/10.1016/j.ssci.2014.04.003

Akyuz E, Celik M (2014b) Utilisation of cognitive map in modelling human error in marine accident analysis and prevention. Saf Sci 70:19–28. https://doi.org/10.1016/j.ssci.2014.05.004

Altuntaş C, Tuna O (2013) Greening logistics centers: the evolution of industrial buying criteria towards green. Asian Shipp Logist 29(1):59–80. https://doi.org/10.1016/j.ajsl.2013.05.004

Antão P, Calderón M, Puig M, Michail A, Wooldridge C, Darbra RM (2016) Identification of occupational health, safety, security (OHSS) and environmental performance indicators in port areas. Saf Sci 85:266–275

Bergantino AS, Musso E, Porcelli F (2013) Port management performance and contextual variables: which relationship? Methodological and empirical issues. Res Transp Bus Manag 8:39–49. https://doi.org/10.1016/j.rtbm.2013.07.002

Bichou K, Gray R (2004) A logistics and supply chain management approach to port performance measurement. Marit Policy Manag 31(1):47–67. https://doi.org/10.1080/0308883032000174454

BIMCO (2015) The global supply and demand for seafarers in 2015. Manpower report, In

Bouejla A, Chaze X, Guarnieri F, Napoli A (2014) A Bayesian network to manage risks of maritime piracy against offshore oil fields. Saf Sci 68(Supplement C):222–230. https://doi.org/10.1016/j.ssci.2014.04.010

Chen L, Zhang D, Ma X, Wang L, Li S, Wu Z, Pan G (2016) Container port performance measurement and comparison leveraging ship GPS traces and maritime open data. IEEE Trans Intell Transp Syst 17(5):1227–1242. https://doi.org/10.1109/TITS.2015.2498409

Chhetri P, Jayatilleke GB, Gekara VO, Manzoni A, Corbitt B (2016) Container terminal operations simulator (CTOS)—simulating the impact of extreme weather events on port operation. Eur J Transp Infrastruct Res 16(1):195–213

Chiu R-H, Lirn T-C, Li C-Y, Lu B-Y, Shang K-C (2011) An evaluation of free trade port zone in Taiwan. Asian Shipp Logist 27(3):423–445. https://doi.org/10.1016/S2092-5212(11)80020-9

Chu S, Cui Y, Liu N (2017) The path towards sustainable energy. Nat Mater 16(1):16–22

Coast Guard, United States (2017) Steam ship El Faro sinking and loss of the vessel with 33 persons missing and presumed deceased northeast of acklins and crooked island, Bahamas on October 1, 2015. In Marine Board's Report. U.S. Department of Homeland Security

Cruz MR, Ferreira JJM, Azevedo SM (2012) Operational performance and physical capacity of Iberian seaport: a multi-criteria analysis. In Studies in Fuzziness and Soft Computing, 449–63

Da Cruz RP, Ferreira JJM, Azevedo SG (2013) Logistics resources in seaport performance: multi-criteria analysis. Marit Policy Manag 40(6):588–613. https://doi.org/10.1080/03088839.2013.777979

De Langen PW, Sharypova K (2013) Intermodal connectivity as a port performance indicator. Res Transp Bus Manag 8:97–102. https://doi.org/10.1016/j.rtbm.2013.06.003

Delandshere G, Petrosky AR (1998) Assessment of complex performances: limitations of key measurement assumptions. Educ Res 27(2):14–24. https://doi.org/10.2307/1176194

Dong Y, Frangopol DM, Sabatino S (2016) A decision support system for mission-based ship routing considering multiple performance criteria. Reliab Eng Syst Saf 150:190–201. https://doi.org/10.1016/j.ress.2016.02.002

Douma A, Schutten M, Schuur P (2009) Waiting profiles: an efficient protocol for enabling distributed planning of container barge rotations along terminals in the port of Rotterdam. Transp Res Part C: Emerg Technol 17(2):133–148. https://doi.org/10.1016/j.trc.2008.06.003

Downing SM (2003) Validity: on the meaningful interpretation of assessment data. Med Educ 37(9):830–837

Ernstsen J, Nazir S, Røed BK (2017) Human reliability analysis of a pilotage operation. Saf Sea Transp:293–300. https://doi.org/10.1201/9781315099088-51

Ernstsen J, Nazir S, Røed BK, Manca D (2016) Systemising performance indicators in the assessment of complex sociotechnical systems. Chem Eng Trans 53:187–192. https://doi.org/10.3303/CET1653032

Fabiano B, Currò F, Reverberi AP, Pastorino R (2010) Port safety and the container revolution: a statistical study on human factor and occupational accidents over the long period. Saf Sci 48(8):980–990. https://doi.org/10.1016/j.ssci.2009.08.007

Grabowski M, You Z, Song H, Wang H, Merrick JRW (2010) Sailing on friday: developing the link between safety culture and performance in safety-critical systems. IEEE Trans Syst Man Cybern Part A Syst Humans 40(2):263–284. https://doi.org/10.1109/TSMCA.2009.2035300

Heij C, Knapp S (2012) Evaluation of safety and environmental risk at individual ship and company level. Transp Res Part D: Transp Environ 17(3):228–236. https://doi.org/10.1016/j.trd.2011.12.003

Himes AH (2007) Performance indicators in MPA management: using questionnaires to analyze stakeholder preferences. Ocean Coast Manag 50(5–6):329–351. https://doi.org/10.1016/j.ocecoaman.2006.09.005

Hinkin TR, Bruce Tracey J, Enz CA (1997) Scale construction: developing reliable and valid measurement instruments. J Hosp Tour Res 21(1):100–120

Hu Z-H (2015) Multi-objective genetic algorithm for berth allocation problem considering daytime preference. Comput Ind Eng 89:2–14. https://doi.org/10.1016/j.cie.2015.04.035

Iannone R, Miranda S, Prisco L, Riemma S, Sarno D (2016) Proposal for a flexible discrete event simulation model for assessing the daily operation decisions in a Ro–Ro terminal. Simul Model Pract Theory 61:28–46. https://doi.org/10.1016/j.simpat.2015.11.005

Kandakoglu A, Celik M, Akgun I (2009) A multi-methodological approach for shipping registry selection in maritime transportation industry. Math Comput Model 49(3–4):586–597. https://doi.org/10.1016/j.mcm.2008.09.001

Kim TS (2015) The revealed competitiveness of major ports in the East Asian region: an additive market share analysis. Asian Shipp Logist 31(4):429–435. https://doi.org/10.1016/j.ajsl.2016.01.001

Kim T-e, Nazir S, Øvergård KI (2016) A STAMP-based causal analysis of the Korean Sewol ferry accident. Saf Sci 83:93–101

Kluge A (2014) The acquisition of knowledge and skills for taskwork and teamwork to control complex technical systems: a cognitive and macroergonomics perspective: Springer

Lam JSL (2015) Designing a sustainable maritime supply chain: a hybrid QFD–ANP approach. Transp Res E-Log 78:70–81

Leriche D, Oudani M, Cabani A, Hoblos G, Mouzna J, Boukachour J, El Hilali Alaoui A (2015) Simulating new logistics system of Le Havre Port. IFAC-PapersOnLine 48(3):418–423. https://doi.org/10.1016/j.ifacol.2015.06.117

Leveson N (2011) Engineering a safer world: systems thinking applied to safety. MIT Press, Cambridge

Li J, Jiang B (2014) Cooperation performance evaluation between seaport and dry port; case of Qingdao port and Xi'an port*. Int J e-Navig Marit Econ 1:99–109. https://doi.org/10.1016/j.enavi.2014.12.009

Lin K-L, Jhan H-T, Ting K-H, Lin C-L, Liu W-H (2014) Using indicators to evaluate the Taiwanese distant-water fishery-policy performance. Ocean Coast Manag 96:29–41. https://doi.org/10.1016/j.ocecoaman.2014.04.028

Lu C-S, Tseng P-H (2012) Identifying crucial safety assessment criteria for passenger ferry services. Saf Sci 50(7):1462–1471. https://doi.org/10.1016/j.ssci.2012.01.019

Lun YHV (2011) Green management practices and firm performance: a case of container terminal operations. Resour Conserv Recycl 55(6):559–566. https://doi.org/10.1016/j.resconrec.2010.12.001

Manca D, Nazir S, Colombo S, Kluge A (2014) Procedure for automated assessment of industrial operators. Chem Eng Trans 36:391–396

Marlow PB, Paixão Casaca AC (2003) Measuring lean ports performance. Int J Transp Manag 1(4):189–202. https://doi.org/10.1016/j.ijtm.2003.12.002

Mavin TJ, Roth W-M (2014) A holistic view of cockpit performance: an analysis of the assessment discourse of flight examiners. Int J Aviat Psychol 24(3):210–227. https://doi.org/10.1080/10508414.2014.918434

Monteiro JGR (2010) Measuring productivity and efficiency of major ports of India. Econ Polit Wkly 45(26/27):325–331

Moon DSH, Woo JK (2014) The impact of port operations on efficient ship operation from both economic and environmental perspectives. Marit Policy Manag 41(5):444–461. https://doi.org/10.1080/03088839.2014.931607

Naderpour M, Nazir S, Jie L (2015) The role of situation awareness in accidents of large-scale technological systems. Process Saf Environ Prot 97:13–24. https://doi.org/10.1016/j.psep.2015.06.002

Nazir S, Kluge A, Manca D (2014) Automation in process industry: cure or curse? How can training improve operator's performance. Comput Aided Chem Eng. https://doi.org/10.1016/b978-0-444-63456-6.50149-6

Nazir S, Øvergård KI, Yang Z (2015) Towards effective training for process and maritime industries. Procedia Manuf 3:1519–1526. https://doi.org/10.1016/j.promfg.2015.07.409

Okstad E, Jersin E, Tinmannsvik RK (2012) Accident investigation in the Norwegian petroleum industry—common features and future challenges. Saf Sci 50(6):1408–1414. https://doi.org/10.1016/j.ssci.2010.12.012

Pak JY, Yeo GT, Oh SW, Yang Z (2015) Port safety evaluation from a captain's perspective: the Korean experience. Saf Sci 72:172–181. https://doi.org/10.1016/j.ssci.2014.09.007

Park N-k, Yoon D-g, Park S-k (2014) Port capacity evaluation formula for general cargo. Asian Shipp Logist 30(2):175–192. https://doi.org/10.1016/j.ajsl.2014.09.003

Peris-Mora E, Orejas JMD, Subirats A, Ibáñez S, Alvarez P (2005) Development of a system of indicators for sustainable port management. Mar Pollut Bull 50(12):1649–1660. https://doi.org/10.1016/j.marpolbul.2005.06.048

Prajogo D, Tang AKY, Lai K-h (2012) Do firms get what they want from ISO 14001 adoption?: an Australian perspective. J Clean Prod 33:117–126. https://doi.org/10.1016/j.jclepro.2012.04.019

Rødseth, ØJ, Perera LP, Mo B (2016) Big data in shipping - challenges and opportunities. In 15th International conference on computer and IT Applications in the Maritime Industries - COMPIT '16. Schriftenreihe Schiffbau, Hamburg, pp 361–373

Roh H-S, Lalwani CS, Naim MM (2007) Modelling a port logistics process using the structured analysis and design technique. Int J Log Res Appl 10(3):283–302

Roussanaly S, Hognes ES, Jakobsen JP (2013a) Multi-criteria analysis of two CO2 transport technologies. Energy Procedia 37:2981–2988. https://doi.org/10.1016/j.egypro.2013.06.184

Roussanaly S, Jakobsen JP, Hognes EH, Brunsvold AL (2013b) Benchmarking of CO2 transport technologies: part I—onshore pipeline and shipping between two onshore areas. Int J Greenhouse Gas Control 19:584–594. https://doi.org/10.1016/j.ijggc.2013.05.031

Roussanaly S, Brunsvold AL, Hognes ES (2014) Benchmarking of CO2 transport technologies: part II—offshore pipeline and shipping to an offshore site. Int J Greenhouse Gas Control 28:283–299. https://doi.org/10.1016/j.ijggc.2014.06.019

Sadovaya E, Thai VV (2015) Impacts of implementation of the effective maritime security management model (EMSMM) on organizational performance of shipping Companies1. Asian Shipp Logist 31(2):195–215. https://doi.org/10.1016/j.ajsl.2015.06.002

Satumanatpan S, Senawongse P, Thansuporn W, Kirkman H (2014) Enhancing management effectiveness of environmental protected areas, Thailand. Ocean Coast Manag 89:1–10. https://doi.org/10.1016/j.ocecoaman.2013.12.001

Saus E-R, Johnsen BH, Eid J, Thayer JF (2012) Who benefits from simulator training: personality and heart rate variability in relation to situation awareness during navigation training. Comput Hum Behav 28(4):1262–1268. https://doi.org/10.1016/j.chb.2012.02.009

Schinas O, Stefanakos CN (2014) Selecting technologies towards compliance with MARPOL Annex VI: The perspective of operators. Transp Res Part D: Transp Environ 28:28–40. https://doi.org/10.1016/j.trd.2013.12.006

Seguí X, Puig M, Quintieri E, Wooldridge C, Darbra RM (2016) New environmental performance baseline for inland ports: a benchmark for the European inland port sector. Environ Sci Pol 58:29–40. https://doi.org/10.1016/j.envsci.2015.12.014

Serebrisky T, Sarriera JM, Suárez-Alemán A, Araya G, Briceño-Garmendía C, Schwartz J (2016) Exploring the drivers of port efficiency in Latin America and the Caribbean. Transp Policy 45:31–45. https://doi.org/10.1016/j.tranpol.2015.09.004

Sleire H, Dale E (2009) The shipping KPI standard. Retrieved from https://www.shipping-kpi.org. Accessed January 2017

Somanathan S, Flynn P, Szymanski J (2009) The northwest passage: a simulation. Transp Res A Policy Pract 43(2):127–135. https://doi.org/10.1016/j.tra.2008.08.001

Song D-P, Dong L, Drake P (2015) Multi-objective optimization for planning liner shipping service with uncertain port times. Transp Res E-Log 84:1–22. https://doi.org/10.1016/j.tre.2015.10.001

Standal D, Utne IB (2011) The hard choices of sustainability. Mar Policy 35(4):519–527. https://doi.org/10.1016/j.marpol.2011.01.001

Stefanidaki E, Lekakou M (2014) Cruise carrying capacity: a conceptual approach. Res Transp Bus Manag 13:43–52. https://doi.org/10.1016/j.rtbm.2014.11.005

Suárez-Alemán A, Sarriera JM, Serebrisky T, Trujillo L (2016) When it comes to container port efficiency, are all developing regions equal? Transp Res A Policy Pract 86:56–77. https://doi.org/10.1016/j.tra.2016.01.018

Sutomo H, Soemardjito J (2012) Assessment model of the port effectiveness and efficiency (case study: western Indonesia region). Procedia - Soc Behavl Sci 43:24–32. https://doi.org/10.1016/j.sbspro.2012.04.074

Talley WK, Ng MW, Marsillac E (2014) Port service chains and port performance evaluation. Transp Res E-Log 69:236–247. https://doi.org/10.1016/j.tre.2014.05.008

Tichavska M, Tovar B (2015a) Environmental cost and eco-efficiency from vessel emissions in Las Palmas Port. Transp Res E-Log 83:126–140. https://doi.org/10.1016/j.tre.2015.09.002

Tichavska M, Tovar B (2015b) Port-city exhaust emission model: an application to cruise and ferry operations in Las Palmas Port. Transp Res A Policy Pract 78:347–360. https://doi.org/10.1016/j.tra.2015.05.021

Tongzon JL (2009) Port choice and freight forwarders. Transp Res E-Log 45(1):186–195. https://doi.org/10.1016/j.tre.2008.02.004

Utne IB (2007a) Acceptable sustainability in the fishing fleet. Mar Policy 32(3):475–482. https://doi.org/10.1016/j.marpol.2007.09.017

Utne IB (2007b) System evaluation of sustainability in the Norwegian cod-fisheries. Mar Policy 31(4):390–401. https://doi.org/10.1016/j.marpol.2006.10.006

Vaghi C, Lucietti L (2016) Costs and benefits of speeding up reporting formalities in maritime transport. Transp Res Procedia 14:213–222. https://doi.org/10.1016/j.trpro.2016.05.057

Wang Y, Yeo GT, Ng AKY (2014) Choosing optimal bunkering ports for liner shipping companies: a hybrid fuzzy-Delphi-TOPSIS approach. Transp Policy 35:358–365. https://doi.org/10.1016/j.tranpol.2014.04.009

Wiegmans B, Witte P, Spit T (2015) Characteristics of European inland ports: a statistical analysis of inland waterway port development in Dutch municipalities. Transp Res A Policy Pract 78:566–577. https://doi.org/10.1016/j.tra.2015.07.004

Wiggins G (1993) Assessment to improve performance, not just monitor it: assessment reform in the social sciences. Soc Sci Rec 30(2):5–12

Woo SH, Pettit S, Beresford AKC (2011) Port evolution and performance in changing logistics environments. Marit Econ Logist 13(3):250–277. https://doi.org/10.1057/mel.2011.12

Yang Z, Ng AKY, Wang J (2014) A new risk quantification approach in port facility security assessment. Transp Res A Policy Pract 59:72–90. https://doi.org/10.1016/j.tra.2013.10.025