Validation of a Behavioral Marker System for Rating Cadet's Non-Technical Skills

Citation for the original published paper (version of record):
Placido da Conceicao, V., Mendes, J., Teodoro, M. et al (2019). Validation of a Behavioral Marker System for Rating Cadet's Non-Technical Skills. TransNav, 13(1): 89-96.
http://dx.doi.org/10.12716/1001.13.01.08

(article starts on next page)

TRANSNAV
http://www.transnav.eu
the International Journal
on Marine Navigation
and Safety of Sea Transportation
Volume 13
Number 1
March 2019
DOI: 10.12716/1001.13.01.08

Validation of a Behavioral Marker System for Rating Cadet's Non-Technical Skills

V.P. da Conceição
Chalmers University of Technology, Gothenburg, Sweden
CINAV, Portuguese Naval Academy, Lisbon, Portugal
Instituto Dom Luiz, Faculdade de Ciências, Lisbon, Portugal

J. B. Mendes
CINAV, Portuguese Naval Academy, Lisbon, Portugal

M. F. Teodoro
CINAV, Portuguese Naval Academy, Lisbon, Portugal
CEMAT, Instituto Superior Técnico, Lisbon, Portugal

J. Dahlman
Chalmers University of Technology, Gothenburg, Sweden
VTI, Gothenburg, Sweden

ABSTRACT: Simulator-based training assumes a very important role in the maritime domain, particularly in the education of Officers Of the Watch (OOW). In the Portuguese Navy, most of the cadet's skills as future OOW rely on the success of this training. Beyond theory and technical training, the development of non-technical skills is a key factor for obtaining officers capable of identifying and solving problems. To optimize the training and development of non-technical skills, using the Portuguese Naval Academy Simulator, a previously designed Behavioral Marker System model was further assessed. The revised model, which comprises new parameters such as the effectiveness of the task, was validated through a set of simulated sessions, where 11 instructors and 48 students participated. After each session, data was collected with questionnaires and focus group discussion, focusing on the quality and usability of the model and on the design of the scenario. The results show that the revised model, positively addresses the limitations found on the previous version, and it has received encouraging feedback from both instructors and cadets. This new model is now under implementation in all the Naval Academy course programs, and future research aims to digitalize the behavior markers.

1 INTRODUCTION

The main goal of this research was to validate a nontechnical skill evaluation model, developed by Conceição et al. (2017), prior to its adoption by the Portuguese Naval Academy. Following the recommendation made in previous work (Conceição et al. 2017), the model was further developed, proposing a revised model (MODACEN) that was assessed by navigation instructors, with a series of trials in bridge simulators. This revision includes a comparative analysis between the MODACEN and the Functional leadership model implemented at the Portuguese Naval Academy (Bué et al. 2015; Pacheco et al. 2015). The validation of the MODACEN model was carried out through the implementation of bridge simulator training sessions and questionnaires to assess the quality of the model.

The MODACEN model contributes to the education of the Portuguese naval cadets, enabling more objective and structured evaluations of the training in simulator sessions. This study also approaches questions directed toward the efficient use of the Bridge Navigation Simulator, concerning the manning requirements.

1.1 *Non-technical Skills*

Non-technical skills (NTS) are define by Flin, O'Connor and Crichton (2008, p.1) as the cognitive, social and personal resource skills that complement technical skills and contribute to safe and efficient task performance. From another perspective, in the context of vocational education and training (VET) there is a focus on the development of transversal competences (Ceitil 2016) or transferable generic competences (Deist & Winterton 2005) that share many of the characteristics of NTS. In Maritime Education and Training (MET) domain, these concepts represent an important complementary approach of the NTS framework. Notwithstanding the consolidation of Bridge Resource Management (BRM) courses in MET programs, focused in the development human behavioral and non-technical skills, we still need to better validate the effectiveness of such skills and training in safe maritime navigation (Barnett et al. 2006, p.9; Oltedal & Lützhöft 2018, p.86; Salas et al. 2006, p.410; O'Connor 2011, p.372). Still, the association of NTS with safe and efficient performance is widely discussed in the human-factor literature (Grech et al. 2008; Oltedal & Lützhöft 2018; Hetherington et al. 2006). On the other hand, the reduction of navigation risks does not rely only on the bridge team performance, since other organizational issues must also be tackled (Manuel 2011, p.34; Hetherington et al. 2006).

It is also relevant to note that both technical and non-technical skills are inextricably intertwined, since they cannot be separated (Flin et al. 2008; Barnett et al. 2006, p.5). Fjeld, Tved and Oltedal (2018) reviewed how the NTS have been applied in the ship bridge domain. After analyzing nineteen studies, they identified five NTS: situation awareness (SA), decision-making (DM), workload management, communication, and leadership. However, they suggest that bridge officers' NTS are not sufficiently explored, calling for a detailed taxonomy and better understanding of the interconnections between the cognitive and interpersonal skills.

1.2 *Behavior markers*

How can we verify that a given individual has the required skills? Considering that competencies are in the first instance behaviors (Ceitil 2016), in order to classify a competency, we need a set of indicators or Behavioral Markers. These Indicators are observable nontechnical behaviors, in teams or individuals, that contribute to superior or inferior performance within a given working domain (Flin & Martin 2001; Klampfer et al. 2001, p.10). Klampfer (2001) suggested essential characteristics of good markers: only behaviors operationalized through observable indicators should be considered as the target of evaluation; with causal relationship to performance outcome; described in domain specific language; using simple phraseology; and describing clear concepts. Ceitil (2016) also raises the question of the standardization of evaluation, to assure the objectivity of the evaluation, implying that each competence should have more than one verification element / indicator. Formal assessment using behavioral rating systems started with the assessment of the effectiveness of Crew Recourse Management (CRM) training of flight deck crew, and by the end of 90's they spread across several domains (Flin et al. 2008). Apart from the prototype behavioral marker for naval officers' NTS designed by O'Connor and Long (2011) and Conceição *et al.* (2017) for behavioral markers of naval cadets training in simulator, few developments are found with a firm employment of a marking scheme for the Bridge Resource Management (BRM) framework (Fjeld et al. 2018; Conceição et al. 2017).

1.3 *Training in Simulators*

Barnett *et al.* (2002) consider simulation a tool to solve problems associated with risk and crisis management, as well as for optimization of navigation team's resources. Elashkar (2016) claims that 58% of the skills associated with resource management of a ship bridge could be improved through simulation and training in simulator. However, several issues need to be addressed, such as the extension of skill transfer from training environment to the working domain, the effective assessment of the NTS, the association with safe performance, the design of the simulator training program (Ward, Hancock, & Williams in Ericsson et al. 2006; Pekcan et al. 2005). Simulators are designed in order to reproducing parts of a real situation allowing its user to practice and demonstrate skills in a controlled environment ensuring integration into the physical context of the task (Hontvedt 2015, p.6).

Studies indicate that individual's performance in a simulation context is a viable source to predict the performance of the same individual in a real context (Mjeldea et al. 2016). However, Sellberg (2016) adds that despite the recognized capabilities of simulators in the learning process, the organization and conduction of the training process is more important than the capabilities of the simulator itself. The need to develop and establish adequate training models to enable and optimize the use of simulators is fundamental to an effective training (Sellberg 2016).

From an educational perspective, using a simulator entails teaching technical skills, developing coordination and teamwork, and evaluating individual and team performances (Hontvedt 2015, p.5). Therefore, the simulator should be properly adapted to the educational context, *i.e.* the level of realism of the simulator must be weighted according to the training objectives and being too close to reality can prevent the identification and / or evaluation of a specific component. According to Sellberg (2017), a higher degree of realism requires more structured training, enabling a close connection between training goals and the particularities of the individuals' performance during the sessions.

The implementation of a set of clear and coherent evaluation criteria that allow the quantification of a subject's performance, covering the whole range of solutions that can be adopted to solve a problem, is a serious challenge. In this sense, Sampson *et al.* (2011) alerts to the problem that instructors in the area of maritime navigation have little knowledge and reveals great uncertainty in the area of assessment skills in simulated environment. Salas *et al.* (2002) had already discussed this misperception that subject matter experts should drive the design of training, suggesting that they should work in collaboration with teaching/learning experts.

Elashkar (2016) proposes that evaluation in simulators should comprise the following elements:

- Collection of evidence that the student has outlined a plan for solving the problem;
- Observation of how the student uses the resources at his disposal;
- Monitoring how the objectives of the exercise are transmitted to the other participants of the session;
- Identification of how tasks are delegated;
- Determine the quality of compliance with the proposed plan;

Despite the challenges linked with the evaluation in simulated environments, the process presents some clear advantages over traditional written tests, providing greater evidence of the student's understanding of the contents evaluated, eliminating factors such as the possibility of plagiarism, copying or memorization of content, that are common failures in traditional processes.

1.4 *The use of simulator in the Portuguese Navy*

The Portuguese Navy has a network of simulators that allows it to manage and guide training in simulator sessions in different training facilities. The Naval Tactical Training Center (NTTC) is the fleet training and evaluation organization and runs a set of simulators that cover both tactical and navigation domains, concurrently or not. The bridge simulator is a KONGSBERG, POLARIS certified Full Mission Bridge Simulator, class A, according to the requirements of the International Maritime Organization (IMO). It comprises seven bridges, four located at the Naval Academy and three at the NTTC, all working in the same network.

1.5 *MODACEN Model*

This model is a revision of the model presented by Conceição *et al.* (2017). NTS framework has five categories: Leadership, Situational Awareness, Communications, Team work and Decision Making. Each category is assessed by three behavioral markers. The main difference is the inclusion of a measure of effectiveness. Other differences fall within the scope of presentation, clarity and usefulness of the form used by the evaluator. The form is subdivided into 3 distinct parts:

- Header: identifies the evaluator, the trainee(s), the place where the evaluation is carried out and the scope in which it is evaluated;
- NTS framework: the core of the evaluation model, it encompasses the different NTS and respective behavioral markers. It contains a section for comments and a section for the evaluation of each indicator between 1 and 5 or not observable.
- Efficiency level: to measure the success achieved by the trainee in the context of the exercise;

1.6 *The Functional Leadership Model*

The Naval Academy runs a skill development program based on the functional leadership framework. Functional leadership is a process of leadership centered on monitoring and intervention in the execution of a task by a team, in which the competencies performed by the leader are the link between the task execution, teamwork, and the associated performance (Santos et al. 2008). The leader assumes a fundamental role in the functional context, being responsible for diagnosing problems, generating / planning solutions and implementing those solutions in the context of the task.

The Functional Leadership Model is operationalized several times over the 5 years of the cadets training program. The symbiosis between the Functional Leadership Model and the one proposed is clearly noticeable, since it is evident that both: 1) privilege the behavior of the leader as a factor of success and main influence in team`s performance, at the same time encouraging the distribution of activities and the autonomy of the team members; 2) establish a close link between the quality of leadership skills and the effectiveness associated with the task; 3) implement cycles of information processing and decision making; 4) promote an environment of inter-assistance, cooperation and communication between the whole team.

The main differences are at the level of operationalization of the use of the model. MODACEN seeks to respond to a set of requirements aimed at the Officer of the Watch's (OOW) performance. On the other hand, the Functional Leadership Model is applied in a wider domain and directed to less complex and structured tasks. The Functional Leadership Model collects elements for evaluation from three moments: briefing, action and debriefing. On the other hand, the MODACEN model is only designed for, and teste during the action itself. It does not collect information from the briefing and debriefing.

2 METHODOLOGY

Data was collected by questionnaires presented to simulator instructors, to assess the prototype model after a practical use in test sessions. Given the reduced number of the participants, the statistical confidence level was set to 90% or $\alpha = 0.1$ for the tests performed.

The MODACEN tests were conducted at Naval Academy Bridge simulator, over a set of 6 sessions, between February and April 2018, involving:

1. 11 participants with instructor duties;
2. 48 trainees assessed with MODACEN model;
3. 3 staff elements to support the operation and control of the Simulator;
4. 2 researchers involved in supervision and conduction of the sessions;

The participation of instructors and trainees was voluntary, after invitation by e-mail. All received an information leaflet describing the goal and procedures for the trails.

Considering the scope of the evaluator functions (instructors), the selection criteria were:

- Naval Officers, specialized in navigation: choice based on the whole set of experience obtained throughout their career, associated with the duties of a navigator, usually responsible for the training of the bridge team and for the certification of OOW onboard naval vessels;

– Naval officers having carried out training using bridge simulators: selection based on the skills acquired and put into practice in the use of navigation simulators for training purposes, preparation of exercises and understanding of the dynamic factors involved;
– Naval officers with specific training in organizational behavior, leadership and teamwork: the competencies gained by these elements, particularly at the level of non-technical skills, plays a key role in their choice to participate in the tests. Their understanding of the theoretical framework of non-technical knowledge leads to a more effective and reasonable evaluation;
– Instructors from the Nautical School leading STCW Pilot Master courses and conducting training in simulators. Their perspectives do not consider any specific naval doctrine applied in navigation onboard naval vessels. In addition, their academic and professional curriculum in the maritime field and training enriches the evaluation, clarifying the adaptability of the conceptual model outside the scope of Naval education.

Trainees were recruited from the Naval Academy cadet's corps, for each session, organized as described in Table 1.

Table 1. Duty distribution among cadets.

Duty	Cadets grade	graduation program
OOW	4th year	Line Officer
ECDIS operator	4th and 3rd year	Line Officer
ARPA operator	4th and 3rd year	Line Officer
Chart work	2nd year	any
Helm and telegraphs	1st year	any
Lookout	1st year	any

The duty arrangement considers the courses' programs and levels of skills expected for each course and year. The aim in structuring teams with cadets from different years is to eliminate factors such as comfort and familiarization between individuals. The fact that a team is composed of elements with different knowledge, requires a constant readjustment, by the team leader, which highlights the management capacity of the team.

Before each trial session, a briefing was held for the participating trainees, to explain the procedure, requirements, and the goals and scenario that were set for the session. Afterwards, a presentation of about 25 minutes was given to the instructors to: clarify the scope of the research, present the developed model, briefly explain the test scenario and the functions inherent to the role of evaluator. At the same time, during this period, the cadets were adapting to the environment of the ship's bridge and organizing themselves. During this period the cadets would revise the information sheet, where general instructions for the series would be presented.

After completing the test series, cadets and evaluators, went to the bridge simulator briefing room for debriefing, collecting contributions and other comments. The evaluators then responded to the evaluation questionnaires.

The series began with cadets and evaluators in their respective bridge, bridge simulator operators and the session supervisor in the control room. The start signal was given by the supervisor to all bridges. At this time the cadet with OOW functions had access to the 1st block of instructions. The series had a duration of 60 minutes, during which the cadets had to deal with 5 specific cases, without any interference of evaluators.

Despite some prior validation tests of the scenario, the conduction of the session underwent successive changes based on the comments and recommendations gathered during the briefings and debriefing of the first sessions. These changes focused particularly on how the supervisor interacted with the evaluated participants (cadets), corrections and improvements in the presentation of the briefing, and adjustment in the times for insertion of the cases played in the series.

3 RESULTS

A total of eleven instructors, male (n = 8) and female (n = 3), took part in the test, aged between 30 and 54 years, 8 from the Portuguese Navy and 3 from the Nautical School (all STCW qualified masters and lecturers). The majority hold a MSc degree (n = 7), two with a PhD degree and other two with a bachelor's degree. All have experience in evaluating sessions in simulated environment and at least 7 have more than 3 years of experience in the field of training. Courses and training in the domain of NTS and instruction in simulated environment are sparse. However, in general, respondents feel they have a good or very good preparation for observer functions (n = 9 / 81,8%). One of the most salient factors in the sample, is the evidence that the level of training obtained by the respondents for teaching and assessment functions is clearly limited in the areas of non-technical skills and interaction with simulation systems. Thus, for the purposes of statistical analysis, the experience associated with each individual was considered as the main correlation factor.

The information analyzed in this study derives from the fifteen questionnaires filled out by the eleven elements of the sample. Three of the eleven elements participated in more than one session and, therefore, given that they had the possibility of having contact with the model again, they were submitted to a new questionnaire in order to measure how their opinion varied according to the corrections made to the model and the conduction of the session.

3.1 *The scenario*

Sections II and IV of the questionnaires assess the respondents' opinion on how a simulated environment assessment session should be conducted and their views on the scenario played during the test sessions. All the respondents gave a very high level of important to the existence of briefing and debriefing. The recommended number of trainees per session, according to the respondents, can be seen in Table 2. The ideal number of trainees, on average, is around five elements, with a minimum of three and a maximum of seven.

Table 2. Number of trainees per session (Q 16).

	N	Min	Max	$\bar{x}$	s
Ideal number of trainees	12	2	6	5,08	1,084
Min. number of trainees	14	1	5	3,36	1,082
Max. number of trainees	14	4	8	6,79	1,311

To inquire for any correlation between the number of trainees and the respondents' experience, the following non-parametric tests were made: Mann-Whitney, Kolmogorov-Smirnov and Spearman's correlation. The results show that the data are not relatable (p> 0,1) so there is nothing to conclude about the relationship between the number of trainees per session and the experience of the respondents.

In the context of the place where the respondent would be to evaluate a training session, the majority (66,7%) prefers to be in the cubicle where the action takes place, whereas 26,7% prefer to be in the control room with access to audio and video.

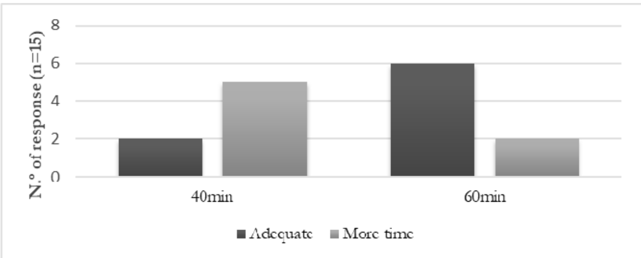

Figure 1. Evaluation of the session duration, Sig=0,072 (qui-square).

Based on the played scenario, when asked about the duration of each session, 53% of the respondents recognize that it was adequate for the ongoing evaluation and the remaining 47% think that the sessions should be longer. However, after the first three sessions, the duration was extended to 60 minutes based on the input's given by the participants at the debriefing. The implications of this change in the opinion of the respondents is illustrated in Figure 1.

When questioned about the exercise technical difficulty, the respondents believed that it was adequate to the objectives of the session. 93.3% consider that increasing the number of students of different years might bring the exercise closer to a real scenario.

3.2 *THE MODACEN*

The third section of the questionnaire assess not only the validity of the model according to the criteria under evaluation, but also the quality of the medium used (forms and information leaflet). We verify (Figure 2) that the respondents characterized the model as a good functionality tool.

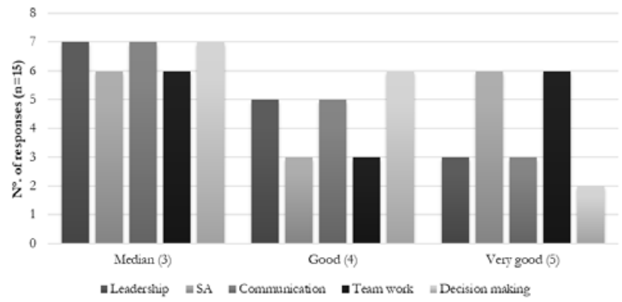

Figure 2. Evaluation of the functionality of the form.

One of the core questions in the questionnaire is to assess respondents' opinion on the benefit of implementing the model to provide more objective assessments of non-technical competencies. The results are clearly satisfactory, since the totality of respondents recognizes the model as an added value.

When asked about the degree of importance of non-technical skills presented in the context of OOW training, the average of the degree of importance associated with each competency is around 4.5, on a scale between one and five, revealing that respondents feel that these competencies are indeed important (summary results presented in Table 3).

Table 3. Statistical results on the NTS's importance (Q 25).

NTS	N	$\bar{x}$	s	Min	Max
Leadership	15	4,53	,516	4	5
SA	15	4,47	,640	3	5
Communication	15	4,53	,516	4	5
Team work	15	4,67	,488	4	5
Decision making	15	4,53	,516	4	5

When asked about the ability of the model to evaluate the NTS required for OWW, looking at Figure 3 we see that the respondents attribute a positive balance to the model, being clear that it leaves room for improvements given the number of respondents who assess the model only as median.

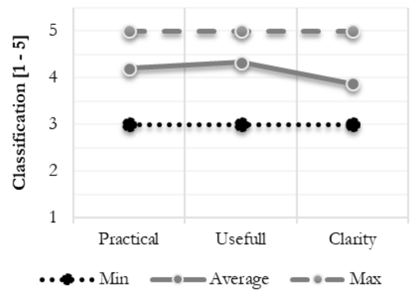

Figure 3. Classification of MODACEN as a tool (Q 26).

By subjecting the variable (ability of the model to evaluate the NTS- Q 26) to the Wilcoxon test according to the importance of NTS (Q 25), assuming H0: there is no relation between the use of MODACEN as an evaluation tool and the importance of the evaluated NTS, it is verified that there is a relationship between the variables (p <0.1), taking into account that the importance given to each NTS tends to be greater than MODACEN's capacity to evaluate it.

Regarding the importance assigned to each of the indicators for the NTS evaluation, the results show that respondents attribute on average, on a scale of importance from one to five, a score of four for all indicators. It should also be noted that, several respondents consider that the different indicators of NTS should be differently weighted, as shown in Figure 4.

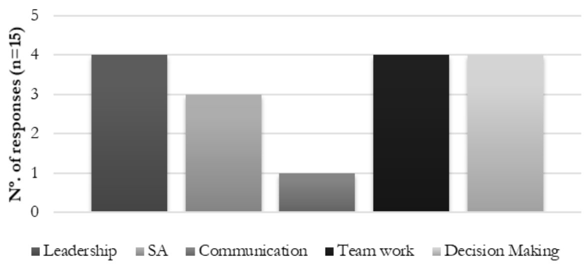

Figure 4. Different weights for the NTS' indicators

For the *Leadership* indicators, the respondents who defend the allocation of different weights for each variable (26.7%), consider that the indicator "*Set intentions and goals*" should be twice as important as the others. Regarding *Teamwork*, 26.7% of the respondents emphasized that there is a close dependence between the behavioral markers, proposing that the indicator "*Coordinates the tasks of the team should*" should overweight the others. *Decision-making* is marked by the opinion of some respondents who state that the weights given to each marker lacks the specificity of the situation and a more detailed characterization of the problem. The skills *Situational Awareness* and *Communication*, since they represent values lower than 20% of the sample were not considered.

As for the difficulty experienced by the respondents in observing the different indicators, results show that on average the most difficult indicators to observe are: "*Assess the capabilities and corrects procedures*" (3,20), "*Promotes a constructive environment for Communications*" (3,47) and "*Assess and verifies the consequences of the decisions and actions*" (3,47). The most difficult NTS to observe is *Decision Making* (3.69) and on average the respondents feel a difficulty in the order of 3,86 which corresponds to a positive weighted average.

3.3 *Coherence of the evaluations*

Each session was evaluated by at least two evaluators. Thus, the relationship between the evaluations assigned by each evaluator to the session leader was analyzed to assess whether the use of the model reflects the desired coherence. All the six session we analyzed on a case-by-case basis, based on the assessment made to the 15 behavioral markers.

Table 4 summarizes the average variation of the different behavioral markers. The results show that in the set of six sessions (seven evaluation series) the most distinct indicators among the evaluators are:

- Establish and control standards;
- Keeps a continuous, clear and effective flow of information,
- Assess the capabilities and corrects procedures.

Table 4. Average variation of the different behavioural markers

NTS / Behavioural marker	s	Associated variation
Leadership		
Takes the initiative	,606	<1
Set intentions and goals	,707	1
Establish and control standards	,825	] 1, 2 [
Situational Awareness		
Monitor and reports changes of situations	,354	<1
Collects external information	,424	<1
Identifies potential danger or problems	,589	<1
Communication		
Shares information	,505	<1
Keeps a continuous, clear and effective flow of information	,943	] 1, 2 [
Promotes a constructive environment for communications	,471	<1
Teamwork		
Considers all the elements of the team	,707	1
Coordinates the tasks of the team	,303	<1
Assess the capabilities and corrects procedures	,825	] 1, 2 [
Decision Making		
Establishes alternative lines of action	,707	1
Assess and verifies the consequences of the decisions and actions	,471	<1
Considers and shares with the others, the risks of the different lines of action	,589	<1
Efficiency assessment	,404	<1

The others reveal an acceptably low level of variation for an assessment characterized by subjectivity and the difficulty of quantifying non-technical skills. This consistency in behavioral markers also extends to the evaluation attributed by the evaluators to the effectiveness of the team.

A set of hypotheses were drawn to demystify the subjectivity of the evaluation and to perceive which indicators contribute to an evaluation that does not depend only on the evaluator.

Table 5. Spearman correlation between the synthetic index of preparation of the trainees and the behavioral markers

NTS / Behavioural marker	correlation value	*p*
Leadership		
Takes the initiative	-,097	,741
Set intentions and goals	,398	,159
Establish and control standards	,231	,448
Situational Awareness		
Monitor and reports changes of situations	,319	,288
Collects external information	,197	,540
Identifies potential danger or problems	,162	,596
Communications		
Shares information	-	1
Keeps a continuous, clear and effective flow of information	,199	,514
Promotes a constructive environment for communications	,380	,200
Team work		
Considers all the elements of the team	,149	,610
Coordinates the tasks of the team	,093	,753
Assess the capabilities and corrects procedures	,232	,445
Decision making		
Establishes alternative lines of action	,313	,298
Assess and verifies the consequences of the decisions and actions	,441	,132
Considers and shares with the others, the risks of the different lines of action	,679	,011

The first hypothesis tries to determine the relationship between the training of the trainees in the domains of Collision Regulation, language and operation of the bridge equipment and the evaluation obtained for each one of the indicators. To facilitate the process a synthetic index was created that encompasses the three domains of the preparation of the trainees and the variables were then submitted to the Spearman non-parametric correlation test. The results are presented in Table 5 (H0: there is no relation between the variables). We may see that only one of the indicators (*Considers and shares with the others, the risks of the different lines of action*) has the required level of significance (the null hypothesis does not hold, p <0.1), resulting in a correlation of approximately 68% with growth in the same direction.

Another hypothesis reflects the relationship between the difficulty of observing the indicators and the difficulty of evaluating NTS with the trainee assessment made by the evaluators. Similarly, to the previous hypothesis, a synthetic index was created that associates, through the mean, the difficulty felt by the respondents to observe the different NTS followed by the non-parametric Spearman correlation test (H0: there is no relation between the variables). It is verified that only one data group corresponds has the required degree of confidence (*Establishes alternative lines of action*), with a correlation factor in the order of 60%.

Proceeding in the same way, we tried to verify the relationship between the evaluation of the trainees in the different indicators and the difficulty of observation associated with each. From Spearman's correlation results, two groups of data respond positively to the level of desired significance, *Collects external information* and *Assess the capabilities and corrects procedures*, with correlation levels of 50%, the second with opposite directions of growth.

Table 6. Relationship between indicators evaluation and team effectiveness, for p<0.1

NTS Behavioral marker	correlation value	p
Leadership		
Takes the initiative	,635	,011
Set intentions and goals	,622	,013
Establish and control standards	,755	,001
Situational Awareness		
Identifies potential danger or problems	,552	,056
Communications		
Shares information	,683	,005
Keeps a continuous, clear and effective flow of information	,794	,001
Promotes a constructive environment for communications	,888	,000
Team work		
Coordinates the tasks of the team	,688	,006
Assess the capabilities and corrects procedures	,643	,013
Decision making		
Considers and shares with the others, the risks of the different lines of action	,624	,017

The relationship between the evaluations attributed to each indicator and the evaluation resulting from the effectiveness of the team was also verified. Spearman's non-parametric correlation test (H0: no relation between variables) was used once again with 10 of the 15 indicators within the required level of significance. From the results presented in Table 6, we see that all the indicators present a correlation level higher than 50%, with the indicator *Promotes a constructive environment for Communications* with the highest level of correlation (88%). and the mean of the indicators shows a correlation of 81%.

4 CONCLUSION

We found that despite the limitations of the model, it is relevant and performs well. In terms of the use of MODACEN, the positive results of the questionnaires and comments of the respondents on the benefit, applicability and importance of the model for the training are significant.

As an evaluation tool, we identify some gaps in the clarification of the behavioral markers, namely their interconnection making the individualization of competencies unworkable. Adding the evaluation of effectiveness also allows to establish a link between the non-technical and technical component, considered particularly beneficial by the respondents and by the Functional Leadership Model.

We must account not only for the subjective nature of this type of evaluation, but also for the design of the exercise that must be structured so as not to compromise the adaptation of the cadets to the complexity of the exercise. The model relies on the presence of evaluators close to the action. This assumption implies that there is always at least one evaluator for each bridge and at least one operator in the control room. Respondents clearly prefer to be in direct contact with the action. This entails more personnel and time for the training in simulator. Given the subjectivity and complexity of the evaluation, the exercise must have at least one hour. Thus, using this model for individual assessments requires significant rise in the staff / student ratio. At the end, we consider that the model has a great potential to respond to the educational needs of the Naval Academy, particularly at the level of OOW training, with a concrete targeted process for cadets' non-technical skills.

REFERENCES

Barnett, M., Gatfield, D. & Habberley, J., 2002. Shipboard crisis management: A Case Study. In *RINA, Royal Institution of Naval Architects International Conference - Human Factors in Ship Design, Safety and Operation*. London, United Kingdom, pp. 131–145.

Barnett, M., Gatfield, D. & Pekcan, C., 2006. Non-technical skills: the vital ingredient in world maritime technology? In *Proceedings of the International Conference on World Maritime Technology*. London: Institute of Marine engineering, Science and technology.

Bué, I.M.G., Lopes, F.C. & Semedo, Á., 2015. The Use of the Portuguese Naval Academy Navigation Simulator in Developing Team Leadership Skills. *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, (June), pp.83–88.

Ceitil, M., 2016. *Gestão e Desenvolvimento de Competências* 2ª Edição. M. Robalo, ed., Lisboa: Edições Sílabo.

Conceição, V.P. da et al., 2017. Development of a Behavioural Marker System for Rating Cadet ' s Non-Technical Skills. *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, 11(2), pp.255–262.

Deist, F.D. Le & Winterton, J., 2005. What Is Competence? *Human Resource Development International*, 8(1), pp.27–46.

Elashkar, M.A., 2016. the Use of Simulation Techniques in the Development of Non-Technical Skills for Marine Officers. *International Journal of General Engineering and Technology (IJGET)*, 5(5), pp.19–26.

Ericsson, A.K. et al. eds., 2006. *The Cambridge Handbook of Expertise and Expert Performance*, Cambridge University Press.

Fjeld, G.P., Tvedt, S.D. & Oltedal, H., 2018. Bridge officers' non-technical skills: a literature review. *WMU Journal of Maritime Affairs*.

Flin, R., O'Connor, P. & Crichton, M., 2008. *Safety at the sharp end: a guide to non-technical skills*, Aldershot England: Ashgate.

Flin, R.H. & Martin, L., 2001. Behavioral Markers for Crew Resource Management: A Review of Current Practice. *International Journal of Aviation Psychology*, 11(1), pp.95–118.

Grech, M.R., Horberry, T.J. & Koester, T., 2008. *Human Factors in the Maritime Domain*, London: CRC Press.

Hetherington, C., Flin, R. & Mearns, K., 2006. Safety in shipping: The human element. *Journal of Safety Research*, 37(4), pp.401–411.

Hontvedt, M., 2015. Professional vision in simulated environments - Examining professional maritime pilots' performance of work tasks in a full-mission ship simulator. *Learning, Culture and Social Interaction*, 7(November 2015), pp.71–84.

Klampfer, B. et al., 2001. Enhancing Performance in High Risk Environments, Recommendations for the Use of Behavioural Markers. In *Behavioral Markers Workshop, GIHRE*. Zürich, Switzerland.: Gottlieb Daimler and Karl Benz Foundation.

Manuel, M.E., 2011. *Maritime Risk and Organizational Learning*, Farnham, Surrey: Ashgate.

Mjeldea, F.V. et al., 2016. Military teams – A demand for resilience. *Work*, 54, pp.283–294.

O'Connor, P., 2011. Assessing the Effectiveness of Bridge Resource Management Training. *The International Journal of Aviation Psychology*, 21(4), pp.357–374.

O'Connor, P. & Long, W.M., 2011. The development of a prototype behavioral marker system for US Navy officers of the deck. *Safety Science*, 49(10), pp.1381–1387.

Oltedal, H.A. & Lützhöft, M., 2018. *Managing Maritime Safety*, Oxon: Routledge.

Pacheco, J., Caetano, A. & Tavares, S.M., 2015. Is training leaders in functional leadership a useful tool for improving the performance of leadership functions and team effectiveness? *The Leadership Quarterly*, 26(3), pp.470–484.

Pekcan, C., Gatfield, D. & Barnett, M., 2005. Content and Context: Understanding the complexity of human behaviour in ship operation. In *RINA, Royal Institution of Naval Architects International Conference - Human Factors in Ship Design, Safety and Operation, 23-24 February*. London, United Kingdom, pp. 99–107.

Salas, E. et al., 2006. Does Crew Resource Management Training Work? An Update, an Extension, and Some Critical Needs. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(2), pp.392–412.

Sampson, H., Gekara, V. & Bloor, M., 2011. Water-tight or sinking? A consideration of the standards of the contemporary assessment practices underpinning seafarer licence examinations and their implications for employers. *Maritime Policy & Management*, 38(1), pp.81–92.

Santos, J., Caetano, A. & Jesuíno, C., 2008. As competências funcionais dos líderes e a eficácia das equipas. *Revista Portuguesa e Brasileira de Gestão*, 7, pp.22–33.

Sellberg, C., 2017. From briefing, through scenario, to debriefing: the maritime instructor's work during simulator-based training. *Cognition, Technology and Work*, (November), pp.1–14.

Sellberg, C., 2016. Simulators in bridge operations training and assessment: a systematic review and qualitative synthesis. *WMU Journal of Maritime Affairs*, (November), pp.1–17