Tampere University of Applied Sciences

ASSESSMENT OF SOFT SKILLS IN INLAND NAVIGATION SIMULATOR TRAINING

Angelique Bosker

MASTER'S THESIS
April 2021

Educational Leadership

ABSTRACT

Tampereen ammattikorkeakoulu
Tampere University of Applied Sciences
Master in Educational Leadership

Bosker, Angelique:
ASSESSMENT OF SOFT SKILLS IN INLAND NAVIGATION SIMULATOR TRAINING

Master's thesis 72 pages, of which appendices 4 pages
April 2021

This thesis is about the best way of assessing and grading soft skills of officers and captains for inland cruise shipping. This research resulted in a plan of approach for assessment and grading of soft skills in inland shipping. A qualitative study was conducted for this research.

The data were collected from interviews and a survey. The conceptual framework resulted in a key theme list for the interview and survey questions. The respondents were a mix of the three stakeholders ((1) the inland cruise company, (2) external training provider, and (3) trainees). Independent experts were added as a fourth group. The data were collected from the respondents via 8 individual interviews and 7 completed surveys.

The theoretical section explored different assessment and grading theories that could be applicable for inland shipping and the target group. A mix of various pieces of literature was combined to develop a conceptual framework.

The findings indicate the importance of transparency and objectivity in assessments. The preparation phase is seen as crucial for the successful implementation and execution of the assessment. Realistic scenarios including briefing and debriefing are of utmost importance and should be based on the experience and knowledge of qualified designers and teachers, followed by the execution of the assessment by the judgment of two qualified assessors. Self-assessment of soft skills was supported by the trainees to be implemented. An important finding is the advice to use a specific rubric. The rubric should consist of the assessment criteria with 5 levels with a grading system of 1 to 5. Requirements for assessment criteria are brief, clear, observable, and written in understandable language.

In conclusion, the three stakeholders involved in this research and the independent experts agreed on key items like scenario choices, briefing, debriefing, peer-to-peer response, and transparency.
A suggestion for future research is to observe the assessment and grading of soft skills in practice.

Keywords: assessment, grading, soft skills, maritime, inland navigation

CONTENT

ABBREVIATIONS AND TERMS

CEO	Chief Executive Officer
EU	European Union
IMO	International Maritime Organization
OECD	Organisation for Economic Cooperation and Development
STCW	Standards of Training, Certification and Watchkeeping for Seafarers
TAMK	Tampere University of Applied Sciences

1 INTRODUCTION

1.1 Background

Over the last 50 years, the ship structure and reliability of ship systems especially with a focus on safety has improved considerably to reduce casualties and increase efficiency and productivity. As a result of these improvements, ships are highly reliable these days. But ship structure and system reliability are only a small part of the maritime safety equation. The key factor in the maritime cluster is people. Human error is the most prominent cause of casualty situations in shipping. Research (Hanzu-Pazara et al., 2008) shows us that in the maritime sector about 75 – 96% of casualties are caused, at least partly, by some form of human error.

In busy waters and near port entrances the navigation is most complex and the officers and captain on board have the heaviest workload and responsibility. They have to effectively use all the equipment and resources available and function as a team to handle the work required to complete a safe transit often including the ship preparations for discharge or loading activities in the port. Stress tolerance, decision-making skills, communication with different cultures and nationalities on board and via telecom with other vessels and coastguard are some examples of skills that are very important for working on board a vessel. Within the maritime cluster, the importance of these skills is becoming more widely recognized. This fits the development and the importance of lifelong learning because of our rapidly changing society and technologies. This trend is also visible in the maritime industry. Especially the changing technologies as a result of innovations and stricter rules and regulations require training courses.

For training and education, most shipping companies work together with external training providers. These external training providers offer several (tailor-made) training courses and educational programs. Simulators and

computer-based training are frequently used in education and training in the maritime industry. They allow training skills, competencies and attitudes in an interactive way. Compared with on board training or training in the real environment it is risk-free, timesaving and safer (Sellberg, 2017). Training based on experience in an emergency or bad weather situation doesn't happen very often in reality, but it is important to prepare officers for these kinds of situations. Simulators offer an unique opportunity for tailor-made training, needed to prepare staff for dangerous situations and to prepare them for the right group behavior and decision processes in all relevant situations which occur or could occur in their work.

In the maritime industry, there are no guidelines for identifying non-technical skills on board vessels and how to assess them (Elashkar, 2016). What is the best way to assess and grade experienced, adult people with the purpose to learn from it instead of punishing or blaming? How do we assess and grade skills like stress tolerance, leadership, decision-making and communication?

1.2 Objective

The objective of the research is to advise about the best way to assess and grade soft skills in inland shipping. This includes a plan of approach and implement these criteria based on an adequate structured and sufficient simulator environment. The need for plans and guidelines is confirmed by the systematic literature review of simulators in maritime education and training of Sellberg (2017). He stated that "more studies are needed to provide guidelines for simulator-based assessments of competency to ensure validity and reliability of the assessment methods". There is no agreed or standard assessment and grading method used for grading and assessing soft skills in inland shipping.

In the process of the development and execution of training courses and assessments, there are three important stakeholders involved. Because of the different interests of the three most important stakeholders, the challenge is the assessment and grading of the training courses.

The three most important stakeholders identified for this thesis are:

- **The inland shipping company**. Important reasons for the shipping company to invest in human resources and in training courses are safety of cargo and/or passengers, their personnel, property, the binding of the personnel because of the shortage of officers, the image of the company, insurance premiums, etc. For personal development plans, selection and promotion goals, and for justification of the company's investment in training courses and request for next year's budget, the company needs evidence and results of the training courses. On the other hand, because of the shortage of officers, retention of qualified personnel is crucial so it is important that they feel safe, comfortable and happy during and at the end of the training course. Assessment and grading of the training courses should not scare them off. Assessing and grading training courses can also help to improve the training courses, design them more tailor-made and/or defining the skills gaps of their personnel. In refreshment training courses, assessing and grading enables the company to analyze the progress in results of training courses during a certain period and to analyze the developments in (soft and hard) skills and competencies. All the above-mentioned reasons make a supported assessment and grading of the training courses crucial for the (inland shipping) company and therefore they requested an assessment and grading system.
- **The external training provider**: companies ask external training providers to execute specific training courses. The interest of the external training provider is a long-term relationship with her customer, the shipping company in this case. This is crucial for the external training provider because of the need for the return of investment in the course development, specific knowledge about the inland water sector and

expensive learning infrastructure (simulators) and for covering the operational costs.

- The **trainees** are adults and many of them have years of experience working on board a vessel. A captain or officer in inland shipping operates quite individually in their function while during their hours of duty day and night, they are on the navigation bridge of the vessel alone for 4 to 6 hours and are not controlled or managed all day. Decision-making at a high level under pressure has to be executed at a relatively young age starting many times at 24 years. It is important that this assessment and grading is executed in a safe learning environment where they feel comfortable and where they are not afraid to make mistakes. Based on the training results and assessment of soft skills a personal development plan can be developed. This is also in the interest of the trainees and they are enabled to improve themselves. On the other hand, this can create pressure for improvement during the same refresher training in the near future.

1.3 Research questions

In this thesis, the main research question is:

How to assess and grade soft skills that meet the transparency requests, interests and needs of the three main stakeholders in inland navigation?

To answer this main research question, six sub-questions are defined:

A. What is the best environment for adult learners for learning and assessment?
B. What are soft skills?
C. What are methods to assess soft skills?
D. What are the criteria for the assessment and grading of soft skills?

E. How to implement the criteria to become an integral part of the training process?

F. What is the impact and value of this (soft skills) assessment and grading criteria to the maritime industry?

1.4 Context

The traditional maritime sector consists according to a very broad description by Policy Research (2008): Shipping, naval construction, naval equipment, maritime services, maritime ports, recreational sailing, offshore, navy, inland shipping, naval repair and others. According to this definition, inland shipping is part of the broader maritime sector. Inland waterway transportation is defined by the OECD (2002) as 'floating craft designed for the carriage of goods or public transportation of passengers by navigable inland waterways.

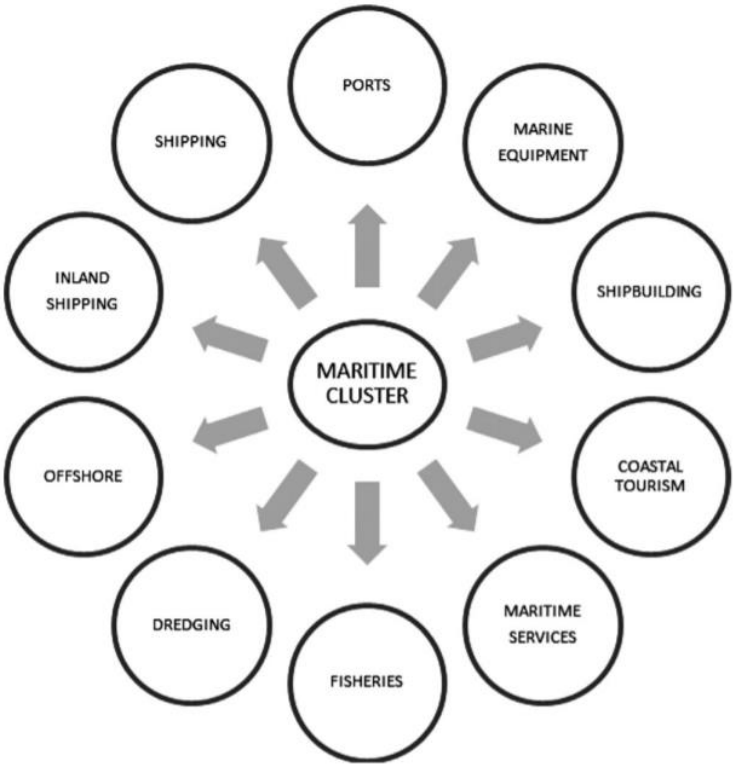

FIGURE 1. Inland shipping as part of the broader maritime cluster. (Pinto, Cruz & Combe, 2015)

Erceg (2018) describes inland waterway transportation as “navigation on rivers, lakes and canals of a certain depth and width that are settled, marked and open for transportation”. Within the maritime sector, there are many differences between the specific sectors as mentioned in the broad “definition” by PRC. Where maritime shipping is international worldwide covered by treaties and regulations via the International Maritime Organization (IMO), inland transportation is characterized by inadequate legal regulation and non-harmonized rules (Erceg, 2018). In maritime shipping the disaster with the s.s. Titanic gave a sense of urgency and was the start for the development of international rules for the safety of the ship and the goods and people on board. This led to treaties like the Standards of Training, Certification and Watchkeeping for Seafarers (STCW) for maritime shipping was adopted in 1978 and gave a boost to quality and a “level playing field” for maritime (sea transport) education. For inland navigation, the EU Directive on the recognition of professional qualifications in inland navigation was only adopted in 2017, about 40 years later. There are also differences in the level and qualification of education. Maritime shipping education is education on higher professional and bachelor level while the inland water navigation qualifications were most of the time based on experience and some short courses only. The level of education can be determined as lower and middle vocational level. Another difference is that inland shipping is not regulated internationally but more regional. The EU Directive refers to inland navigation in the rivers, estuaries and lakes in Europe. In South-East Asia, the Mekong countries, the Amazon area, the Nile area and so on there are no international treaties on education and proficiency level for inland water transport available.

The lack of international regulation and cooperation in inland navigation leads to a lack of studies and research on inland navigation qualifications, training and assessment, the topic of this thesis. Maritime shipping is in this respect an example for inland navigation. Like the structure of the EU Directive, this is based on the structure of the STCW for maritime shipping. The lack of studies on the needs of standards and qualifications, education, training and

more specific on inland navigation assessment and grading is one of the challenges of the inland shipping company. This leads to the need for this thesis research.

In recent years, also inland water vessels and passenger's vessels become faster and larger. The traffic density on the canals and rivers makes safe handling of ships more and more imminent. And the protection of the environment becomes a standard driver for a more structured and standardized approach of training, certification and inland water transport.

1.5 Scope of research

The scope of this research is on grading and assessment of soft skills in inland shipping. This thesis is also characterized by a specific target group. The focus is on adults since they are the employees of the shipping company and are the trainees of the training course. Another focus point of this thesis is the purpose of assessment. Assessment can be used for different purposes. It can be used as part of the selection process, for example for hiring or promotion, or diagnostic to determine trainees' knowledge and skills before a training. The purpose of the assessment focused on in this research is summative assessment: assessment used to measure the soft skills of the trainee at the end of a training course (Isaac-Savage, 2009).

2 ASSESSMENT AND GRADING THEORIES FOR INLAND SHIPPING

In this chapter, the theoretical framework for this research will be explained. First of all, highlights of the learning environment for adult learners will be exemplified since adults are the target group of this research and assessments are part of the learning process. Although this study is focused on assessment, it should not be forgotten that assessment is part of a bigger picture: the complete learning process and experience. Therefore it is needed to take the learning environment and preference of adults into account. Secondly, assessment and grading theories are discussed. Furthermore, soft skills will be defined and the (various) way of assessing them are discussed based on the current literature. This chapter will conclude with the conceptual framework chosen for this study.

2.1 Learning environment for adult learners

We learn every day, in education but also at work, in daily life and groups. Learning has become not only an important topic in psychology and education, but also in an economic context. Skills of employees are an important parameter of competition (Illeris, 2009). Illeris (2007) defines learning as “any process that in living organisms leads to permanent capacity change and which is not solely due to biological maturation or ageing”. Knowles, Holton & Swanson (2005) define learning as “the process of gaining knowledge and/or expertise”. The focus of learning in theories is on the person in whom this change occurs. Spalding (2014) states that maximum learning occurs in a safe learning environment but where the learner feels uncomfortable, but not feeling pushed in the alarm zone.

The process of learning is influenced by many aspects. Not only the basic conditions which include all the psychological, biological and social conditions but also the internal and external conditions are involved in the

learning process (Illeris, 2009). This means that ‘the input’ of the learning process is not the same as the ‘output’, it is different for every person. There are many factors affecting learning, such as age and experience. Knowles et al. (2005) introduce in the early 1970s the concept that adults and children learn differently, the method and practice of teaching adult learners is a theory known as andragogy. The learning process of adult learners is compared to the learning process of children more two-way learning because the experience of the adults is exchangeable with the teacher’s knowledge. This results in shared authority of the learning process. Furthermore, adults also support formulating the curriculum (Knowles et al., 2005).

By comparing the concepts of pedagogy and andragogy, the differences become clear. Pedagogy is focused on teaching children, where the teacher is responsible for the learning content, method, timing and evaluation. Andragogy is based on the following principles about adult learning (Knowles et al., 2005; Houde, 2006):

- Adults need to know why they need to learn something before undertaking it;
- Learners’ self-concept: responsibility for their own decisions and lives;
- Learners’ (variety in) experience;
- Readiness to learn: adults are living their lives while learning;
- Orientation to learning is life-centred (or task-centred) in their approach to learning;
- Strong internal motivators.

The orientation to learning of adults stated that adults learn more effectively when they are confronted with real-life situations. This orientation together with their experience where people learn from, comes together in the experiential learning theory of Kolb, inspired by Dewey. In Kolb’s definition of learning experience plays a crucial role, according to him learning is “a process whereby knowledge is created through the transformation of

experience" (Kolb, 1984). As Kolb (2014) has pointed out, experiential learning has become an important method for adult learning and personal development. The combination of practice and theory helps in their learning process and the integration of work and education. Key in his theory is the need to combine different learning styles: abstract conceptualization (thinking/explaining), active experimentation (applying/doing), concrete experience (doing/feeling) and reflective observation (examining/watching). Thus, learners learn through experience, reflection, conceptualization and experimentation. The combination of different (interactive) learning styles is important for learning. Furthermore, the emphasis in this learning theory is not on the outcomes but the continuous process of learning.

The approach and design of training courses for the adult trainees of the inland shipping company are based on the andragogy principles. One example is the requirement that the instructor/teacher has sailing experience as well. Because of this experience of both the teacher and the trainees, experiences are exchangeably resulting in interesting discussions and more two-way learning. Also, based on his experience, the teacher can provide real-life situations on board the vessels to connect theory to practical situations and to integrate the training course and working on board of vessels. This results in more effective learning as described above. The variety in experience in experience is important to take into account in both a training course as well as in an assessment. Furthermore, the preference for real-life and task-centred should be included in both training courses as in assessment. What is missing in Kolb's approach is the social context of the interaction with other trainees, the teacher and the environment which is an important prerequisite for effective learning.

Summary:

Adult and children learn differently. Many factors affect the learning process, such as age and experience. The adult learning process is more self-directed, two-way learning and they take responsibility. This process is more effective by the integration of work and real-life situations. In this research these andragogy principles for learning environment should be taken into account for the development of assessments.

2.2 Assessment theories

2.2.1 Assessment definition and types

To distinguish assessment from evaluation, assessment will be defined first. Assessment is defined by Moran (1997, cited in Isaac-Savage, 2009) as “using measurements to describe a learner’s achievement and to make recommendations for additional learning activities”. According to Dietel, Herman, & Knuth (1991), it is important to ensure that assessment supports and does not detract from the quality of education. They argued “while assessment has the potential to improve learning for all students, historically it has acted as a barrier rather than as a bridge to educational opportunity. Assessment has been used to label students and put them in dead-end tracks”. This should be avoided. This is supported by Kobayashi (2005) that assessment is influencing the quality of the training. Reeves (2000) states the importance that providers of assessments must seek to reach optimal levels of alignment, alignment among learning objectives, content, instructional design, instructor expertise, technological affordances and assessment strategies.

2.2.2 Assessment methods

Crowe (2000) describes two assessment methods: assessment in a teacher-directed learning and a self-directed learning. In a teacher-directed learning approach, the teacher sets the predefined objectives, assessments and grade standards or criteria. An advantage of this approach is ease to register, disadvantages are the learner focus is on grades, not all subjects are tested and low-level cognitive skills are emphasized.

To encourage independent learning, learners in the self-directed learning approach are involved in decisions about their learning program. The individual learner takes initiative in the whole design of the learning program: from defining the learning needs to the assessment of the learning outcomes. Commonly used as assessment tools are reflection and self-reflection via self and peer feedback. An advantage is that this method increases learning and encourages self-awareness. According to Crowe (2000), self-directed learning is a relief for adult learners compared to the enforced dependency on the teacher in teacher-directed learning in primary and secondary schools. On the other hand, self-directed learning is more demanding of educators in terms of time, creativity and openness. Also for some learners, it can be difficult to switch to this approach and they can become frustrated.

A combination of self and teacher-directed learning is called a learner-centred program. The focus is on self-directed learning, but the facilitator has the final responsibility for the assessment. The “triangulation” method of Hammond and Collins (1991) results in a comprehensive assessment in collaboratively by the learner self, peers and by facilitator. Both learner and facilitator have responsibilities in the quality of the learning program. This is confirmed by the study of Ramm, Thomson and Jackson (2015) to the experience of (senior) students in peer assessment. Based on this research, key benefits were identified as sharing experiences, self-awareness, recognition of their professional development, more confidence and the improvement in

communication skills as it allows communicating both good and bad news to your peers.

The training courses for the inland shipping company are based on a teacher-directed learning approach. The learner-centred approach or triangulation method is interesting to involve the trainees in the development of the program and assessment. This is important to improve the shared responsibility and authority, which is valued by the adult trainees according to Knowles. Therefore, this research will include the application of self and peer assessment, by adding this subject to the interviews. Another reason to add peer assessment is the improvement of communication skills identified as one of the key benefits of peer assessment. This is crucial since communication is often a cause of human error leading to accidents.

Summary:

Assessment should improve learning, not create a barrier or put labels on trainees. Therefore, the training provider should try to reach optimal levels of alignment with the training course. Learner-centred education results in comprehensive assessment via collaboration between learner, peers and the facilitator/teacher.

2.3 Assessment of skills

2.3.1 What are soft skills?

Hard skills, specific competencies needed to perform a job, with all the tools available are nowadays easier to assess (Cimatti, 2016). An example in the maritime industry is to control the speed of the vessel, both in a simulator as well as on a vessel this is easy to check, control and evaluate compared to

soft skills as communication of leadership. The focus of employers and educators is on how to deal with the lack of soft skills training in education. The importance of soft skills in society has increased. According to Cimatti (2016), soft skills are “strategic to be successful in personal and professional life”. But what skills can be identified as soft skills? Soft skills depend on the context and vary widely. Schulz (2008) focuses in his definition of soft skills on personal qualities and interpersonal skills. Communication, critical thinking, problem-solving abilities and creativity are stated as important soft skills. Although some soft skills relate to personal traits and habits, it is possible to change or improve them. Important in this respect is self-recognition, a deficit has to be identified first to be aware of your soft skills that are needed to improve them (Schulz, 2008). Cimatti (2016) uses the term soft skills for “all the competencies that are not directly connected to a specific task; they are necessary for any position as they mainly refer to the relationships with other people involved in the organization”. In the maritime sector, this playfield is enlarged with the soft skills needed with regards to communication, situational awareness, cultural awareness, leadership and behaviour in connection with the colleagues on board and the outside world of port, pilots, coast guard, vessel traffic control and so on. Soft skills are therefore seen as strategic elements and should get high attention during the recruitment process and working life.

2.3.2 How to assess soft skills?

Knowledge can be assessed in a written examination. But a multiple-choice test is not a good assessment instrument for the assessment of skills or performance. According to Miller (1990) the higher the competence pyramid, the more difficult to assess the competencies objectively. In the SAGE Encyclopedia of Contemporary Early Childhood Education (2021) describes objectivity as “information that is collected through measuring, observing and examining facts” and subjectivity as “information that is based on personal

views, opinion, or value judgments". An assessment is never subjective or objective, but a mix of both. Although the question is if any assessment can ever be objective, the availability of a checklist in soft skill assessment "make the assessment process more objective, valid and reliable" (Boashash & Sucic, 2003). Also Miller (1990) states that the impediments can be reduced by a proper design of the assessment test and scoring instruments.

Mueller (2005) states that traditional assessment is not sufficient nowadays anymore. Because acquiring knowledge and skills is not sufficient, it is necessary to learn and demonstrate when and how to apply the knowledge and skills in real-world situations. Alternatively from traditional assessment, performance or authentic assessment is a "form of assessment in which students are asked to perform real-world tasks that demonstrate meaningful application of essential knowledge and skills".

Summarized, the following attributes distinguish authentic assessment from traditional assessment:

TABLE 1. Traditional and authentic assessment (Mueller, 2020)

Traditional assessment	**Authentic assessment**
Selecting a response	Performing a task
Contrived	Real-life
Recall/recognition	Construction/application
Teacher-structured	Student-structured
Indirect evidence	Direct evidence

Miller (1990) uses a pyramid model to describe four different levels of professional competence and associated assessments in health care education. The first one is "knows", knowledge. By multiple-choice questions

and oral exams, this competency can be assessed objectively. The second one is “knows how”, this is the ability to apply the knowledge, such as solving problems and describing and applying procedures. This is possible to assess by essays and multiple-choice questions based on scenarios or cases. The third level is “shows how”, the performance. More sophisticated assessment methods are needed to assess this type of competency and soft skills, like laboratories and simulations to demonstrate the skills. The fourth level is “does”, action, the behaviour in practice. The first and second layers are both cognitive competencies, the third and fourth layers are defined as behavioural competencies. According to Miller (1990) simulators are the right technical instruments to assess (his third layer) behavioural “show how” competencies.

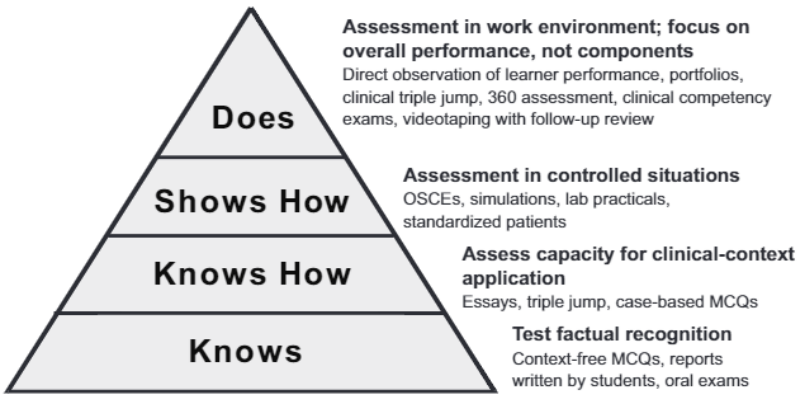

FIGURE 2. Miller’s Pyramid of Professional Competence with examples of assessment techniques used in medical education (Albino et al., 2009)

In connection to this research, soft skills are in the maritime industry of utmost importance. Captains bear a major responsibility for all staff on board and the cargo or passenger on board. Imagine a cruise vessel with thousands of passengers on board, or sailing a cargo vessel with dangerous goods into a port or on the river. It is only a few years ago that the inland shipping companies started with simulator training courses other than the required radar simulator training. After a few years of experience, they see it as an important training tool to improve soft skills and behaviour. Together with an adequate briefing and debriefing of what happened on the simulator and supported by video recording, trainees become aware of their behaviour and

communication as well. The next step, and the objective for this research, is to include an assessment of the soft skills in the training programs.

Summary:

Soft skills are seen as strategic elements in both private and professional life. They depend on the context and it is possible to improve them by training and education. For the assessment of soft skills the traditional way of assessment doesn't fit. Several authors see simulators as a technical tool to show competencies and to assess behavioral competencies.

2.4 The simulator as an assessment tool for soft skills

Based on Miller's Pyramid (1990) and Mueller's authentic assessment (2005), simulators are defined as one of the methods to assess soft skills. Furthermore, simulator-based training is according to different researchers the key factor in training because of financial, environmental, time and safety pressure (Salman, 2013). In high-risk industries like aviation, health care and shipping, simulators are used in education and training because they allow training skills, competencies and attitudes interactive. Compared with on board training or training in the real environment, it is risk-free, time-saving and safer (Sellberg, 2017). Simulators offer an unique opportunity for these types of training, needed to prepare staff for dangerous situations and to prepare them for the right group behaviour and decision processes in these situations which occur or could occur in their work. Also, Kobayashi (2005) stated that navigation and ship handling simulators can be used for objective assessment.

But what is a simulator? Simulators are tools designed to reproduce some aspect of the working environment (Maran & Glavin, 2003). Simulators can be used for different purposes: research, development, testing and education

are some examples. There are different types of simulators: from low fidelity devices like computers, up to simulator-focused approaches to address specific elements (part-task simulators) and high fidelity devices designed to reproduce the work environment in a high fidelity manner (full mission simulator) (Carson-Jackson, 2010). The tool used in training will depend on the learning objective, the target group and the skills to be trained. Simulators can be used for both trainings of technical skills (e.g. ship handling) as well as for soft skills (communication, teamwork, decision making, etc.). The use of a simulator should never be the objective of training. It is an important tool to provide effective training and supports theoretical subjects and assists teachers task in transferring, knowledge, add skills, convince participants and make a clear and transparent examination, assessment and testing possible.

So, a simulator is ‘just’ a tool to provide training and education. Kolb’s learning cycle, as discussed earlier and important in adult learning, is used as a framework for the development and execution of simulator training programs. Simulator training in inland shipping is inspired by the experiential learning cycle of Kolb: theoretical classroom sessions combined with a briefing, concrete and immediate experiences with tailor-made and realistic scenarios in the simulator, observing and reflection during the debriefing session. The influence of the social component and the interactions with others is extremely important in the training courses where the focus is on communication, teamwork, delegation, etc.

In inland shipping low fidelity simulator are used for years. The use of a variety of simulator types including high fidelity simulator is a development of the last decade. Therefore, many inland shipping companies are not aware of all the possibility of a high fidelity simulator. Another difference compared to the aviation industry is that simulator-based training and assessment is more complex because of all the different types of vessels. In the airline industry, only a limited number of aircraft types are on the market (Muirhead,

2006, cited in Pham, 2019) while there are many different types of vessels sailing.

Summary:

Simulator based training courses and assessments enables learners to show their performance in a safe environment with financial, time and environmental advantages. Furthermore, simulators are a great tool to apply the experiential learning cycle ("learning through reflection on doing") of Kolb to create knowledge through the transformation of experience.

2.5 Grading soft skills

Felder & Brent (2010) point out the challenge in assessing soft skills: the grading process. They set two criteria for the grading process of assessing soft skills:

- Reliable: two (or more) assessors grade the score to the given performance, and from the same assessor at different times.
- Fairness: the learner knows the grading criteria and has been adequately instructed in the skills to meet the criteria.

Developing grading checklists or grading rubrics should provide reliability and fairness in the assessment process of soft skills. The checklist consists of grading criteria and the maximum points per criterion, where the grade is the total sum of the points. Also, the rubric describes briefly what each score represents. A weighted sum of the points of each criterion determines the total grade. Grading based on these instruments becomes also much more efficient than the commonly used feedback procedures without a rubric. Another advantage is that it is a tool to show the learners what is expected from them. The grading can become more effective if two assessors grade the learners (Felder & Brent, 2010). For grading, many different scoring scales can be used. According to Mueller (2020), there is no specific number

of levels a rubric should include, it is depending on the task, soft skills and the objective and level of assessment. When using an analytic rubric (a rubric were each criterion is assessed separately), it is advised to start with a small number of levels of performance for each criterion. An example of grading is described by Whitehurst (2016) as the Brookings Soft Skills Report Card. It is a student report card completed by the teacher to measure specific soft skills. Categorized soft skills are scored from 1 (low) to 5 (high).

In the case of this research, no grading systems are used so far. The shipping company and the trainees don't have experience with grading after a training course. Therefore, this will be an important element of the data collection process.

2.6 Assessment and grading frameworks

Based on Mueller's authentic assessment theory (2005), he created the Authentic Assessment Toolbox for the development of authentic assessments standards, tasks, criteria and rubrics and standards for measuring and improving trainee learning based on four steps:

TABLE 2. Authentic Assessment Toolbox. (Mueller, J. 2020.)

Nr	Step	Question	Result
1	Identify the Standards	What should students know and be able to do?	Standards
2	Select an Authentic Task	What indicates students have met these standards?	Authentic tasks
3	Identify the Criteria for the Task	What does good performance on this task look like?	Criteria

4	Create the Rubric	How well did the students perform?	Rubric

For step 3 it is needed to identify the characteristics of good performance of the tasks, the criteria. For each criterion, at least two levels of performance should be identified to sufficiently discriminate among the performance of the learners. In the final step, the development of a rubric shows the combination of the criteria and level of performance for each criterion. A rubric is defined as “a scoring scale used to assess student performance along with a task-specific set of criteria” (Mueller, 2020). Based on a rubric, it is possible to determine the minimum level at which the learners should perform as a benchmark for example promotion or to hire somebody for a certain function. Besides, a rubric gives feedforward to the learner what to improve.

According to Mueller (2005), both trainees and teachers should benefit from the development following the toolbox. By explicitly articulating the standards, assessors or teachers can communicate the goals more clearly to their trainees. Trainees know, based on the defined criteria, what is expected and how to complete the task, therefore trainees are more satisfied and perform better because the goals are clear (Brophy, 1987, cited by Mueller, 2005). This requires well-written criteria that describe observable and measurable behaviour by the instructor.

Kobayashi (2005) defined also a system to develop assessment for maritime training. He stated that the quality of training is influenced by the assessment system for the evaluation of competencies. Step 1 is, like step 1 of the authentic assessment toolbox, to prepare a list with all the assessment items. For example to stay at a safe distance to other vessels. The next step is to define the standard level of competency. Based on the conditions of the simulator scenario and vessel, the quantitative value of the assessment criteria has to be defined. Kobayashi (2005) combines in his second step, steps 3 and 4 of Muellers’ (2005) steps in the Authentic Assessment toolbox.

The third and final step of Kobayashi (2005) is to complete the assessment sheet during the scenario to make sure the assessor remembers exactly enough the actions, not at the end of the session. For the soft skills a scaling score is used of +1 (attained), 0 (lacked) and if applicable/possible -1 (missed). Kobayashi (2005) mentioned also other important aspects. The optimal length of a scenario is between 40 and 60 minutes. The effectiveness of training and assessment is enhanced by a proper debriefing, including showing afterwards the recording of the trainee's behavior during an exercise.

The above two theoretical frameworks for assessment and grading of soft skills are discussed. Muellers' toolbox describes the steps extensively and gives examples. Muellers' toolbox but is a model useable for all kinds of assessment, where the model of Kobayashi is more generally described but more specific for maritime training and assessment. In both models, the importance of the knowledge of the instructor and their uncertainty about the development of (simulator) assessments is missing, as discovered by Sampson et al. (2011, as referred to by Sellberg, 2017). Another aspect missing in these frameworks is the need for familiarization with the simulator: "a high level of familiarization with the equipment and system is required before engaging with simulators for training and assessment purposes" (Maung, 2019). Feedback and continuous review and monitoring of the assessment is another aspect to take into account. Where Kobayashi advises using the scoring +1 (attained) and 0 (lacked), is Muellers' approach open for every grading score that fits the context and situation. However, the questions for each step are very useful and clear in the four-step model of Mueller.

Summary:

A grading rubrics should provide reliability and fairness in the assessment process of soft skills. There are 4 steps describes to define a rubric: identify the standards, select an authentic task, identify the criteria and create the rubric. For grading different systems can be used. Fairness and reliability are identified as important grading criteria.

2.7 Conclusion conceptual framework

In conclusion, it is expected that this combination of the various theoretical models can serve the interests of all three stakeholders. The development of the list with assessment items and the standard level of competencies in the form of a rubric is r for both the teachers of the external training provider and the trainees. Firstly, teachers can communicate clearly about the assessment items and the scoring scale. Secondly, trainees know what is expected and this results in more satisfaction by them and better performance. More satisfaction and better performance by the trainees will lead to more confidence with the inland shipping company to use this assessment system.

Another interesting item to check during the interviews is self and peer assessment. As stated by different authors (Crowe, 2000; Hammond and Collins, 1991) self, peer and facilitator assessment increases learning and encourages self-awareness. Furthermore, this approach to give more responsibility to the trainees fits more to adult learners compared to teacher-directed learning. Also, Ramm et al. (2015) identified key benefits of peer assessment as communication, self-awareness, sharing experience, etc. Kobayashi (2005) and Miller (1990) identify a simulator as an appropriate tool to assess soft skills. Another interview point is the scoring scale for grading the soft skills. Based on his study for maritime assessments, the system of Kobayashi to score an -1, 0 or 1, is proposed but there are also other

systems. Reliable and fairness are identified by Felder & Brent (2010) as grading criteria.

All these expectation will be further investigated during the interviews. Some important questions are: Is the shipping company interested in encouraging self and peer assessment? What does the external training provider think about involving the trainees in the development process? What is the experience with grading systems?

All above considered, leads to the following theories used for the conceptual framework:

TABLE 3. Conceptual framework. (Bosker, 2021)

	Building stone framework	**Author**
Assessment design	Authentic assessment toolbox: 4 steps for assessment development	Mueller (2020)
Self-directed learning approach	Involvement of trainees in the design phase	Crowe (2000)
Assessors	Facilitator, self and peer assessment	Crowe (2000); Hammond & Collins (1991)
Method for assessing soft skills	Use of simulator	Miller (1990); Kobayashi (2005)
Length of scenario	40 – 60 minutes	Kobayashi (2005)
Scoring scale of the performance	+1 (attained), 0 (lacked) and if applicable/possible -1 (missed).	Kobayashi (2005)
Grading criteria	Reliable and fairness	Felder & Brent (2010)

3 METHODOLOGY

The following chapter will show the choice of the methodology for this research: the broad philosophical underpinning. The methodology is used to answer the main research question: How to assess soft skills on a simulator that meets the transparency requests, interests and needs of the three main stakeholders?

3.1 Research design and structure

This research will lead to practical recommendations requested by the shipping company. In social sciences, there are in research tradition two main streams: quantitative and qualitative research. Quantitative research is deductive and focuses on the quantification in the collection and analysis of data (Bryman, 2016). On the other hand, research focused on meanings, interpretation and experiences is called qualitative research (Bryman, 2016). Where the focus of quantitative research is on gathering information from a variety of data, qualitative research is focused on collecting information on one topic or phenomena to enrich the desired understanding. The objective of qualitative research is to explore deeply a specific topic to develop a more person-focused practice (Thomas & Magilvy, 2011).

The methodology is according to Melnikovas (2018) “a general research strategy which delineates the way how research should be undertaken”. It helps to underpin the choices made for the research approach and methods. The research onion of Saunders, Thornhill and Lewis (2019) is a way for methodology construction. It supports organizing the research and to develop the research design. The onion consists of six layers: philosophy, approach to theory development, methodological choice, strategy, time horizon and technique and procedure of data collection and analysis.

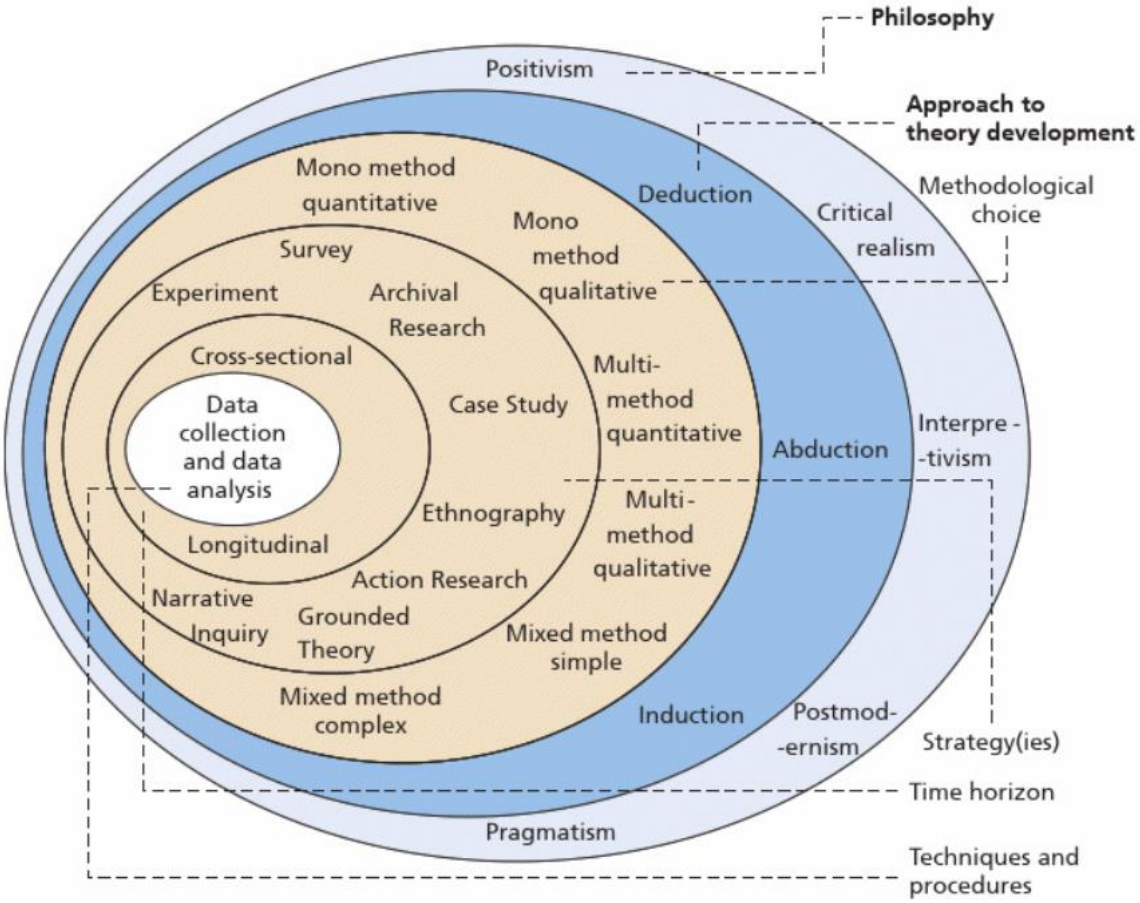

FIGURE 3. Research Onion (Saunders, Thornhill and Lewis, 2019)

3.2 Research philosophy

Research philosophy is a system of beliefs and assumptions about the development of knowledge, in this case, this is addressing a specific problem of the shipping company wherefore new knowledge is developed. There are five major philosophies. There is no one ‘best’ philosophy for business and management research. This thesis, as part of the master program at the Tampere University of **Applied Science**, focuses on finding a practical solution and outcome for a problem. The knowledge of this thesis is needed to enable actions that should be carried out successfully. This approach fits the pragmatism philosophy, where research starts with a problem. In this case, the problem is that at the end of a training course, there is no assessment and grading system and therefore it is not possible to compare the results between the trainees and the training results over the years. The

nature of the reality, the ontology, is complex and rich and the reality is the practical consequences of ideas. In this research the ideas of the stakeholders. As the shipping company, the pragmatists are interested in the practical outcome of this research not in the abstract distinction. The pragmatism philosophy seeks to “improve practice by the application of concepts”. Therefore, the research question is seen as the most important determinant for the research design. This is what this thesis is about: combining the different interests of the stakeholders into an assessment and grading system. In this respect, the socially constructed nature of the reality of interpretivism should not be forgotten. Humans and their meaning and interpretations of reality are different from studying physical phenomena (Saunders et al., 2019). Different people, different organizations, different (organizational) cultures, under different circumstances make different meanings and create and experience different social realities. In this research, the people and their ideas and experiences, the organizations (stakeholders), the context (economic, political, etc.) and the culture of the organizations will determine which assessment and grading system will fit at this moment to the organization (culture), shipping company.

3.3 Approach to theory development

The second layer of the research onion defines the approach to theory development. This layer includes two contrasting approaches: deductive and inductive approach. A third approach used often in business is a combination of the deduction and induction approach: abduction. These approaches can be classified into theory building (induction), theory testing (deduction) and theory modification (abduction). This research was based on the abductive approach. This paragraph will explain the reasons for choosing this approach.

First, this research has started with a problem or surprising fact: there is no assessment and grading at the end of the training courses. Because these systems are missing, it is not possible to monitor the development and improvement of the trainees, to compare the skills of the trainees and to develop more specific tailor-made training courses based on the ‘to be developed’ skills. Based on this problem, plausible theories were collected for finding a solution to this problem. One of the characteristics of an abductive approach is modifying an existing theory. In this case, the theoretical framework showed there is information available in other contexts: numerous studies are available about assessment and grading in general and at primary and secondary schools. But far less in the context of this research: the maritime industry and none for inland shipping. An abductive approach enables to modify of the existing theories. Therefore, existing theories were collected to identify patterns and themes (for example the development process, the assessment and the grading system). Since there was no existing theory found for our problem, theories are mixed. Mixing or modifying existing theories is a characteristic of the abductive approach. This leads to a conceptual framework. Based on this conceptual framework, questions for the interview and survey were defined to collect data and to examine to what extent the mixed existing theories fits the practical problem and solution of this research.

The inductive approach is not possible since there is no assessment and grading system in place yet, so there is nothing to observe. Conducting interviews without collecting the theoretical knowledge first, would also make no sense since the knowledge about assessment and grading available within the company is limited and they need the ‘help’ of existing theory to know which possibilities and systems there are.

3.4 Research design

The research design exists of the three next layers: methodological choice, research strategy and time horizon. It is important to achieve a coherence research design plan.

3.4.1 Methodological choice

The third layer in the research onion of Saunders et al. (2019) is the methodological choice. Therefore it is needed to define if a quantitative or qualitative or mixed research design is followed. To answer the main research questions of this research it is needed to collect the interests and needs of the stakeholders and to allow the stakeholder to share their experiences and preferences. Where quantitative research is used in research about numeric data, as numbers, qualitative is about non-numeric data. Qualitative research is based on meanings expressed through words and resulting meaning derived from words (Saunders et al., 2019). This research was a qualitative study that investigated and answered how the assessment and grading of soft skills can be approached, developed and implemented. The collection of the expectations, needs and interests of the stakeholders were not possible to express in numbers. The objective was to explore deeply the soft skills assessment to develop an assessment tailor-made (more person-focused practice) and practical solution for the shipping company, taking the interests of the other stakeholders (trainees and external training provider) into account.

3.4.2 Research strategy

The fourth layer of the research onion is about the research strategy. The research strategy is also part of the research design. A research strategy is a plan how to answer the research questions. This research is defined as a case study. Case studies try to find an answer on ‘how’ and are defined as an in-depth study into a phenomenon within its real-life setting (Yin, 2018). A case can be a person, organization, an event or many other types of case subject. In this research, the single case was focused on soft skills assessment and grading system that fits the inland shipping company. According to Dubois and Gadde (2002), a case study is the best way to study the interaction between a phenomenon (soft skills assessment and grading system in this case) and the context (within the maritime industry at a specific company). This case originated from practice. The inland shipping company experienced obstacles and felt insecure in the development and implementation of an assessment and grading system for soft skills. Also, the other two stakeholders, the trainees and the external training provider have to deal with their own needs and interests and struggled with the assessment and grading system for soft skills. Therefore, this research answers the practical need of this challenge by identification of critical aspects of an assessment and grading system and this provides practical recommendations. In the current literature, there is a gap concerning taking the various interests and needs into account.

3.4.3 Time horizon

The time horizon of this research is a ‘snapshot’ of the situation at a particular time, also called a cross-sectional study. This research started in August 2020 by defining the problem, collecting theories and defining the research plan. The interviews were conducted in December 2020 and January 2021. The results of the interviews conducted are based on the ideas, experiences

and perspectives of the interviewees employed at the different stakeholders at that moment. New experiences with assessment and grading or new people at the positions within the organizations, will lead to other results. To minimize this influence, there is chosen a fourth group of interviewees: independent simulator and education experts. They are independent experts and therefore seen as more objective and transparent in their ideas, experience and perspectives.

3.5 Data collection and analysis

As in other qualitative research and because of this unique case study, this research used non-probability sampling. This sample included therefore a subjective judgement. Purposive sampling, also known as judgemental sampling, is very often used in case studies when it is needed to select an informative sample. A heterogeneous sampling is chosen of participants with diverse characteristics, experience, needs and interests. This is not only needed to provide a maximum variation possible in the data collection (Saunders et al., 2019) but also to create support and commitment of all three stakeholders. This enables this research to collect the different perspectives on the key themes and to describe and explain them. The sample existed of the three stakeholders (inland shipping company, trainees and the external training provider) and the fourth group of independent experts is added as stated above. To collect the data, in total eight interviews were conducted with the stakeholders of the inland shipping company, external training provider and independent experts. The selected interviewees have also various background: from psychologist involved in promotion assessment, education experts, former CEO and managers of a maritime training and education provider and former captains/master mariners. Because of practical reasons, working on board of vessels, bad internet connection and night shifts, the data collection of the seven trainees was conducted via a survey instead of via interviews as planned. To be able to compare the interests and ideas of the trainees with the other stakeholders, most

questions in the survey were as the questions in the interview (so open questions). During the interviews and surveys, qualitative data is collected. This is primary data: collected by myself without making use of other data collected before. The data is collected in various ways:

- 6 one-to-one interview meetings are held online, via video tools and technology;
- 2 face-to-face interviews were conducted;
- The ideas and experience of the trainees are collected via a survey tool: Survio.

Because of the pandemic, most interviews and so, the data collection process was held online. The duration of the interviews was 1 to 1,5 hour per interview.

The collected data is non-standardized and needs to be classified for analysis. A single data collection technique is used in the form of semi-structured interviews, a so-called mono method qualitative study (Saunders et al., 2019). For collecting qualitative data in-depth and semi-structured interviews were used. Semi-structured interviews are chosen because although they are flexible and they allow the interviewer to diverge to pursue an emergent idea, interest or experience in detail to understand the perception of the stakeholder (Yale University, 2015), the structure is needed to ask the main questions to all stakeholders to be able to compare their answers.

For these in-depth and semi-structured interviews, a list of the themes is developed based on the conceptual framework. Questions related to these themes are used as a guideline for the interviews. There is chosen for semi-structured interviews because of the differences in backgrounds and experiences of the interviewees. An in-depth interview requires the researcher to frame the questions to each participant and interpreting their answers (Saunders et al., 2019). These questions are not asked in the same

order and way, it depends on the interviewee. The key themes list with example questions was sent to the interviewees in advance through email, so they could prepare themselves and think about the themes in advance. The interview progressed spontaneously and not all questions were asked to all interviewees. For example, if the topic was discussed in a previous answer or if the interviewee had no experience or idea about a topic. During the interviews, textual data was collected as notes from the interviews and survey. This key themes list was also used for categorization for the data analysis. It is the basis for the thematic analysis. The answers of the interviewees were ordered per theme. Coding and labelling were used to categorize the answers.

TABLE 4. Interview themes. (Bosker, 2021)

Key principles	-Main challenges of soft skills assessment
	-Requirements
Breadth and balance of assessment design	-Self-directed learning approach
	-Methods and tools
	-Use of simulator
	-Duration scenario
Execution of assessment	-Joint responsibility / Learner-centred approach
	-Rubric
Evaluation of performance	-Grading system
	-Criteria

3.6 Summary

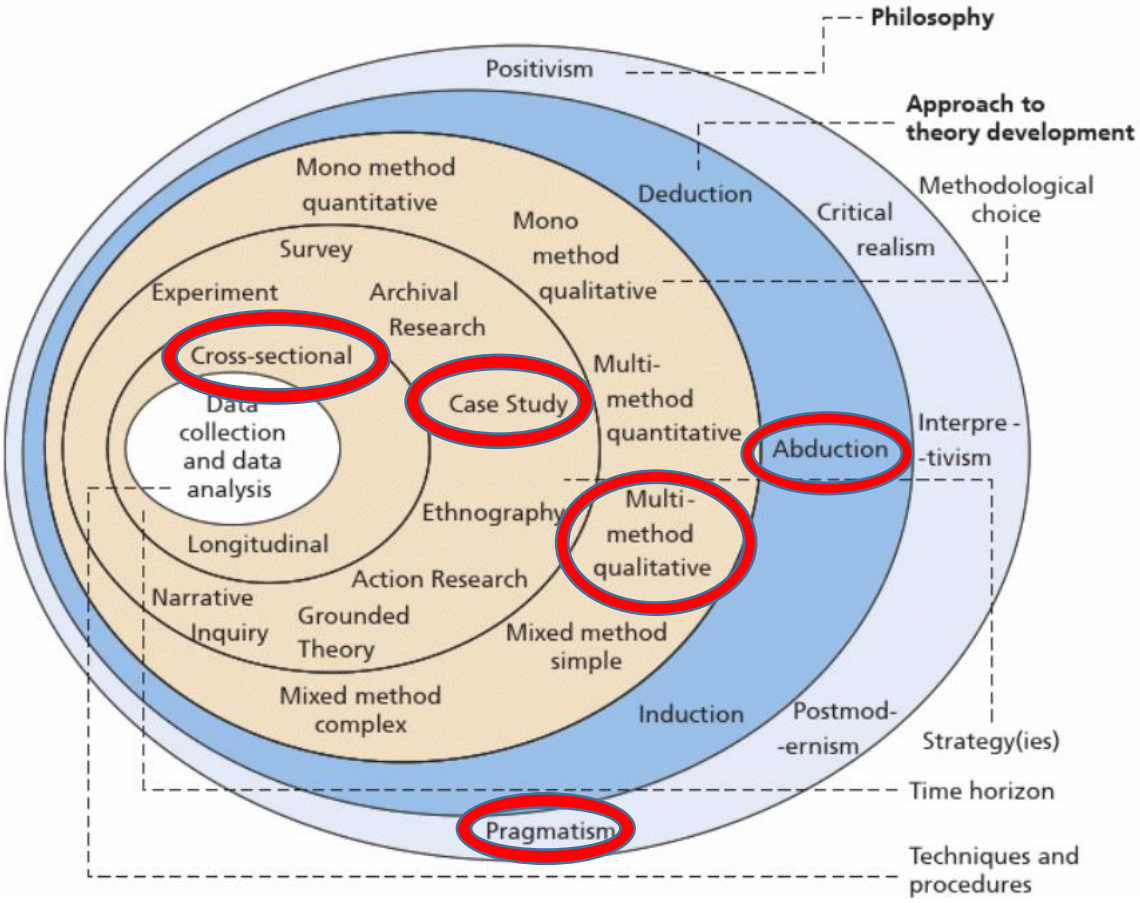

FIGURE 4. Choices research onion (Bosker, 2021)

4 RESEARCH RESULTS

The following chapter will present the research results from both the interviews with the shipping company, the external training provider and the experts, as well as the surveys with the trainees.

4.1 Challenges and requirements of soft skills assessment and grading

The interviewees identified the main challenges of soft skills assessments. Objectivity and unbiased assessment, well-prepared assessments and qualified assessors are identified as main challenges. During the preparation especially the objective of the assessment, what should be assessed and the assessment framework (in Mueller's model step number 1: Standards) was mentioned as important. Another interviewee mentioned the challenge of creating a safe and comfortable environment and personalized approach. While the role of the assessor was not mentioned in the assessment theories described in the literature review, it is confirmed by learning and simulator theories. According to Sellberg (2017), their work of organizing and facilitating the learning activities before, in and after-action (briefing and debriefing) is crucial for meeting the (learning) objectives as well as the professional guidance and feedback.

Also, the culture was mentioned by three interviewees as a challenge, especially the culture in inland navigation. First of all, because inland shipping is a very practical job. Many skills are learned on the job, on board the vessel. Besides the positive experience of learning on board, it becomes a problem if something is learned not correct. As an example of this, learning English from a nonnative English speaking colleague was mentioned during an interview. Especially when competencies and skills are assessed by these

non-qualified colleagues, for example by the same nonnative English speaker.

Another aspect of the culture that influences soft skills development and assessment is that the work in the bridge is individual and independent, it is doing yourself and less teamwork oriented. Therefore, assessment by third parties but also by peers or 360-degree assessments is not common in this sector. This is also different compared to maritime shipping. Training and functioning in onboard of seagoing vessels are more focused on team building, communication and working together (like the training bridge resource management).

Various items were mentioned as crucial requirements for soft skills assessment. Proper familiarization at the start of the assessment was mentioned as a requirement. Since there are many vessel types, equipment manufacturers and bridge layouts, the simulator or simulation used for assessment are not the same as the vessel where participants are sailing on. Therefore time is needed to get used to the equipment, options and vessel manoeuvring, etc. Two interviewees suggest a minimum of 2 scenarios for familiarization. Duration familiarization is 2.5 to 4 hours.

Secondly, preparation time for the developers of the assessment is experienced as a requirement. It is experienced by interviewees who also have experience with the development of the assessment that the preparation time is often forgotten or underestimated. Preparation time is needed to develop real-world scenario’s that fit company-specific situations and company philosophy about learning and development. Also during the preparation, it is needed to define objectives, to describe competencies and to develop scenarios and a clear process of the training and timeslots. Most forgotten is time to test the assessment and scenarios.

Furthermore, the briefing at the start of the assessment to manage the expectations was mentioned as a requirement. Topics as the objective of the assessment, what will happen with the results and who will receive the report should be communicated during the briefing. This requirement is related to another requirement: the quality and number of the assessors. At least two qualified assessors need to be involved in an assessment who can give objective judgements based on the assessment items and criteria. Objectiveness and integrity are crucial for soft skills assessments.

4.2 Breadth and balance: Assessment design

As Crowe (2000) stated that self-directed learning is a relief for adults since it encourages independent learning, the adult trainees could be involved in decisions about their learning program and in the development process. The interviewees are asked if they see opportunities to include participants in the development phase and if they have experiences with it. All interviewees agree that it is a good idea to involve them in the development process. One of them stated that this improves the intrinsic motivation and participation of the trainees. According to another interviewee, it results in more transparency and consistency, which is sometimes missing in bigger organizations. Therefore, involvement in the development process can increase the commitment to the assessment and the organization. However, there are doubts about the practical implementation of it. None of the interviewees has experience with involving students or adult trainees in the development process of an assessment. In practice, a representative group could be asked to join the development process.

Another important note is the difference between assessment at the end of a training course and assessment as part of a selection or promotion. In the development process of an assessment at the end of a training course, it is possible to involve trainees, in a selection or promotion assessment it is not

possible since they are more sensitive and the “surprise level” of unknown factors should be part of these assessment scenarios. Another question asked to the interviewees is appropriate methods and tools that enable assessing and grading soft skills. This is related to step number 2 of Mueller’s toolbox: select an authentic task. An authentic task is characterized by and differs from traditional tasks through:

- performing a task,
- the task replicates challenges faced in the real world,
- application of knowledge
- student structured
- direct evidence

In this sense, simulation scenarios can be seen as an authentic task, but are they also seen by the interviewees as an appropriate tool for soft skills assessment? The interviewees see simulators as the next best tool to assess and grade soft skills, after on board training and assessment. Since it is because of time, safety and practical reasons not possible to simulate extreme weather circumstances, engine breakdown or fire in the engine room and other disasters, simulators are seen as an appropriate tool. Another reason why it is seen as an appropriate tool is that a simulator enables to combine both navigation skills and soft skills, as in real life. An interviewee supports her statement that it can be seen as an appropriate tool because 95% of the trainees recognize him or herself in the report based on the simulator assessment. One of the interviewees says that a simulator is a very good way to assess and grade soft skills, but also warns that it is a very strong tool that can break the best captain. This puts an extra burden on the scenario design and assessor.

The next question related to the development of the assessment is the length of a scenario. The optimal length of the assessment is according to Kobayashi (2005) 40 up to 60 minutes. All interviewees agree with this, however, most of them mentioned that the time for briefing and debriefing

should be calculated as well. Every scenario should include an organized and structured briefing and debriefing for reflection and feedback facilitated by a qualified teacher/assessor. A briefing is an instruction by the instructor of the assignment to the trainees. It is focused on practical information about the upcoming scenario and the objectives (Sellberg, 2018). Practical information can be information about the vessel, the area/location of the vessel and where to sail to. The objectives can also include the assessment items and the way of grading. These instructions are open and straightforward and the specific events and incidents in the scenario are still unknown (Sellberg, 2018). The trainees are asked in the survey to the information they need before the start of the assessment. Most of the trainees ask for very detailed nautical information about the exercise, they mention the location of the vessel, information about the river stretch, weather forecast, traffic, route, etc. Surprisingly, only a few of them mention the purpose or the objective.

During the debriefing, the instructor reconnects the briefing to the scenario and the objectives and/or assessment items. Simulator technology makes it possible to playback the scenario including voices and mimics of the trainees. Therefore all actions are becoming observable and open for feedback, reflection, analysis and discussion. This is described in the literature as very important to integrate knowledge with practical experience (Sellberg, 2018). In practice, the trainees expect during the debriefing a recap of the simulation, feedback and advice and showing different solutions or possibilities to solve the challenges in the scenario.

The duration of a scenario is also important to take into account in a plan of approach. Firstly, because time is money. Companies prefer to send their trainees to a training or an assessment as shortly as possible. In many cases, they prefer one or two days only, but for familiarization requires also some time. Secondly, the planning of a training course or assessment depends on the duration. The total time for a scenario, including a briefing and debriefing, estimated by the interviewees is about 1,5 hour up to 2 hours. From this total

time, 45 minutes up to 75 minutes is needed for playing the scenario. It depends amongst others on the (number of) assessment criteria in the simulator scenario.

The biggest challenges of assessment, and part of the assessment design, is to write down clear, understandable criteria of the rubric. Identifying criteria (step 3 of Mueller's toolbox) is not seen as a problem, the soft skills criteria are clear based on experiences and examples from maritime shipping. The toolbox gives characteristics of good criteria. According to Mueller (2020), criteria should be "clearly stated, brief, observable, statement of behaviour and written in language students understand". All these characteristics were mentioned during various interviews, but applying this is seen as the biggest challenge of assessment design. Related to this, the challenge is to find a balance between the number of items and the level of detail of a rubric. On the one hand, it is important to describe the assessment items very precisely. If the items are not described accurately, the assessment of the items becomes more subjective (for example the trainee communicates well). On the other hand, if the list exists of too many specified items it is impossible to observe, assess and grade all items in a scenario. Therefore, during several interviews examples of communication as an assessment was discussed just like different example and experiences with rubrics.

For example:

- No further breakdown of communication:
    - *Communication* (Does the candidate communicate in a clear, effective and informative manner?)
    - Communicates clear following the standards ("River speak")
- Distinguish internal and external communication
- Distinguish communication in:
    - Two-way communication
    - Thinking out loud
    - Asking questions
    - Sharing ideas

- Communication via VHF(external communication): short, simple and clear

In Mueller's toolbox there are two guidelines included relevant for this study:

- Limit the number of criteria (less than 10): keep it to the essential elements of the task;
- You do not have to assess everything on every task.

4.3 Execution of assessment

After the design of the assessment, it is time for the execution of the assessment. Realistic scenarios including challenges, clear objectives and assessing the soft skills needed to be assessed were some example of what the trainees are expecting from an assessment scenario. These items are were also mentioned by the shipping company, the external training provider and the experts as requirements of an assessment. Two interviewees were not familiar with the term 'rubric' or had no experience with using it. All others were positive about the use of a rubric score scale with the criteria and level of performance.

The next question to the interviewees was related to the types of assessors of the assessment. According to the self-directed learning approach, the joint responsibility should include a combination of self, peer and facilitator assessment. During the interviews, the term assessor is used by the interviewees instead of facilitator (the term used in literature). Although they all agree that it depends on the objective of assessments, the advice of the interviewees to may or may not include self and peer assessment are different. All interviewees agree that peer-assessment should not be used for selection or promotion assessment due to the competition between the trainees. In case of a training assessment, peer assessment can be meaningful, but some of the interviewees have the experience that trainees

are too kind to each other during peer assessment. It is not very neutral and unbiased since you know that your fellow trainees will assess you as well. If peer assessment is included, a clear briefing is needed to prepare the trainees how to assess their peers. Training on how to give feedback can be included as well. Peer assessment can also be done anonymously in writing and scoring but there is no experience with this method.

The experienced assessors stated that they used self-assessment, but not by completing the same rubric or a written form. They used it verbally by asking the trainees questions during the debriefing part: How do you think you did it? What went well? What could be better? What would you do differently next time? By asking these questions to the trainee, the assessor guides the trainee to an accepted, committed and supported judgement about his performance instead of giving him advice and have to tell him what could be better.

The results of the survey completed by the trainees were in line with the answers of the interviewees. In the survey, the trainees scored assessment by the assessor, self-assessment and peer assessment. They could divide 10 points between these three options. Most of the points, more than 51%, were given to assessment by the assessor. Almost 33% of the points were given to self-assessment, and only 16% to peer assessment. What do trainees expect from an assessor? They expected an assessor who is professional, fair and honest. All trainees should be evaluated the same way. Also important is to explain during the debriefing what and why went wrong during the exercise, no judging but advising on how to deal with these kinds of situations to learn from them.

4.4 Evaluation performance

The scoring or grading system was an interesting discussion during the interviews. This is related to the last step, step 4, of Mueller's toolbox: finalizing the rubric with various levels or labels of how well did the trainees perform. Mueller doesn't give a specific number of levels a rubric should possess. Kobayashi (2005) advises to use for assessing the soft skills a scaling score of +1 (attained), 0 (lacked) and if applicable/possible -1 (missed). The interviewees did not agree with this advice. Most interviewees suggested using a scoring scale of 1 (poor) up to 5 (excellent) because in this system you can add the nuances of what is preferred by the trainees. However, some of the interviewees who suggested a score of 1 up to 5 use in practice a scoring scale of 1 up to 3. They stated that it is very difficult to distinguish between a score of 1 and 2 and between 4 and 5. Sometimes 0 is added to scale as the soft skill is not observed during the scenario exercise.

Another grading system used by an interviewee is -1 (poor), 0 (sufficient) en 1 (good). Also – and + are used as grading system instead of numbers, where:

-- = Performance not meeting the requirements at all
- = Insufficient performance
+/- = Adequate performance, but room for improvement
+ = More than sufficient performance
++ = Very good performance

The advice of most interviewees was in line with the opinion of the trainees based on the survey: more than 71% preferred a grading system from 1 up to 5. Two interviewees have experience with and suggest adding 'critical items' or also called 'knock-out criteria' to the system. Critical items are criteria that should be scored at least sufficient because these soft skills are crucial for the execution of the function. If the score on these items is insufficient, the score on the other items is irrelevant because the final result

is insufficient. It is advised to label a maximum of two or three criteria as critical items.

Another important piece of advice of one of the interviewees was to NOT complete the rubric including the grades during the scenario but doing this afterwards. This is in contradiction with Kobayashi (2005), who stated to complete the assessment sheet during the assessment. The interviewee advised to write down all kind the observations during the assessment. One of the pitfalls of an assessor is to judge too fast, for example, if the communication in one situation was very good but later on it can go wrong. Therefore it is better to observe during the scenario and write down your observations first, without a judgement. At the end of the scenario, both assessors should complete the rubric independently. After the completion, both assessors will discuss and combine their scores. This is a very helpful and thorough remark and suggestion. Grading criteria reliability and fairness are seen as most important. Also, validity is mentioned as an important grading criterion: the extent to which an assessment measures what you intend it to measure. Other interviewees mentioned that the boundaries and differences between the various levels should be clear and described in understandable language.

4.5 Results compared to the conceptual framework

The above results from the interviews and survey are not completely in line with the conceptual framework inspired by the theories. This paragraph compares the building stones of the conceptual framework, developed in paragraph 2.7, with the results described in paragraph 4.1 up to 4.4. For the assessment design, the 4 steps of the authentic assessment toolbox were recognized and advised. The involvement of trainees during the design process as recommended by the self-directed learning approach is seen as

too complex by all participants of this study. It is suggested to establish a representative group for commitment and support during the design phase. Assessment by at least two qualified assessors is seen as a requirement. Besides, self-assessment is supported for self-awareness and self-enhancement. Peer assessment is not supported by most participants of this study. The use of a simulator was seen as an appropriate method to assess soft skills in inland shipping. Per scenario, a duration of 45 up to 60 minutes was advised in line with the conceptual framework. For the evaluation of the soft skills not the grading score of -1, 0 and 1 is preferred, but the interviewees and trainees preferred a grading system of 1 up to 5. Validity and clear boundaries between the scores are added to the grading criteria reliable and fair. Table 5 shows an overview of the building stones of the conceptual framework and the study results.

TABLE 5. Results conceptual framework. (Bosker, 2021)

	Building stone framework		**Results**
Assessment design	Authentic assessment toolbox: 4 steps for assessment development	✓	The 4 steps are recognized by the interviewees.
Self-directed learning approach	Involvement of trainees in the design phase		Too complex; a representative group is suggested.
Assessors	Facilitator, self and peer assessment		Facilitator and self-assessment are supported.
Method for assessing soft skills in inland navigation	Use of simulator	✓	The best method to use for soft skills assessment.
Length of scenario	40 – 60 minutes	✓	45-60 minutes for a scenario, including briefing and debriefing 1,5 – 2 hours.
Scoring scale of the performance	+1 (attained), 0 (lacked) and if applicable/possible -1 (missed).		Most interviewees and trainees prefer to use a scoring scale of 1 (poor) up to 5 (excellent).
Grading criteria	Reliable and fair	✓	Reliable and fair are seen as important criteria, validity and clear boundaries/differences are added.

5 PLAN OF APPROACH FOR ASSESSMENT IN INLAND NAVIGATION

The combination of theory, the interviews and the survey results give clear advice for a plan of approach for the design and implementation of an assessment and grading system in education and training for inland navigation. During the interviews and survey, crucial aspects were mentioned that should be taking into account. Some of these aspects were missing in the reviewed theories.

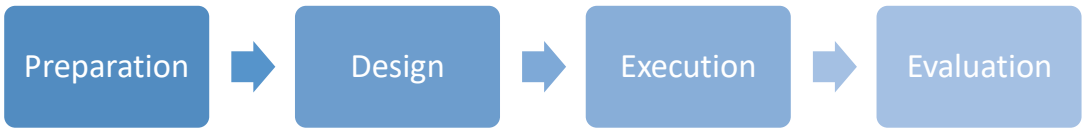

FIGURE 5. Phases of assessment based on interviewees and survey (Bosker, 2021)

Phase 1: preparation

Since there is limited attention in the literature for the importance of the preparation time, it is in accordance with the experience people interviewed very important to take enough preparation time into account. The time needed is according to them underestimated. According to the external training provider, this time is crucial to figure out the customer needs and objectives of the assessment together with the inland shipping company. The lack of international qualifications in inland navigation in combination with very limited experience with training and assessment demands more time to clarify this. After the customer needs and objectives are known, it is needed to define:

- Who will you assess? Define the target group.
- What is the current/starting level?
- What is the number of participants?

- What is the duration of the assessment?
- Is there training prior to the assessment provided?

The inland shipping company and the external training provider should align and discuss these items in depth before going to the next phase. It is advised by the external training provider to discuss in this stage the number of assessors needed during the assessment. Two assessors are needed for the four-eyes principle, but due to the lack of experience in training and assessments, inland shipping companies are not used to the extra costs of an assessor. Before the start of the next phase, the assessment design, it is recommended to discuss who will be involved in the design phase. Besides representatives of the inland shipping company and the external training provider, it is advised to bring together a representative group of trainees and to involve them during the assessment design.

Phase 2: design

The next phase is the design phase of the assessment. As stated by several interviewees, enough time for this design phase is crucial for the execution of the experience, the preparation of the assessors and how the trainees (and assessors) experience the assessment. To start the design first, a schedule for the duration of the assessment, if applicable including the training course, should be determined. If the assessment follows after a training course, the design phase should start with the design of the assessment and after that the training course. Important is to include two familiarization scenarios. If an assessment follows at the end of a training course, a familiarization is not needed when the simulator was included in the training course and there was time to familiarize. The number of trainees should also be taken into account. It is advised to use small groups, depending on the (number of) simulators. This is needed to ensure all trainees have enough time to show their skills. Another advantage is that this helps to create a safe and comfortable environment and personalized approach. The time for briefing and debriefing, in total about 1,5-2 hours per scenarios, should be included in the schedule

as well. For the assessment scenarios, it is needed to have two assessors available.

After this organizational part of the assessment, it is time to identify the standards of the assessment, step 1 of Mueller's toolbox. This step is followed by step 2 and 3 of Mueller's toolbox: determine the authentic task, or in this case the simulator exercise, and determine the criteria, also called assessment items. The simulator exercises should be realistic and include the criteria. Therefore it is needed to write down in detail the exercise to be sure all assessment items are included. During the assessment design and scenario development, it is needed to take the individual and specific culture of the inland navigation into account.

Two main challenges of assessment are to determine the number of criteria and to translate the soft skills into criteria and describe them well. These two challenges are also related to each other. For both challenges, Mueller (2020) gives some advice. He recommends limiting the number of the criteria (maximum 10) and not all criteria have to be included in every scenario. Furthermore, Mueller (2020) gives characteristics of good criteria: it should be clearly stated, brief, observable, statement of behaviour and written in language the trainees understand. During this step is also needed to identify the critical soft skills items. It is advised to determine one or two critical soft skills related to the objective of this assessment and to label them as critical items (knock-out items). After determining the criteria the next step is to create the rubric. A 5 point scoring scale (1 – 5) will be used. Therefore it is needed to describe the criteria for all 5 levels. Finally, the scenario should be tested. If the developers of the assessment are not the same people as the assessors, an adequate briefing for the assessors before the start is crucial.

Phase 3: Execution

The assessment starts with an introduction of the assessment. What is the objective? What is expected from the trainee? What is the focus of this assessment? What will happen with the results/report? All kind of information to ensure the expectation management of the trainees, to give personal attention and to create a safe and comfortable environment.

Every scenario starts with a briefing with the role division between the trainees, the situation of the vessel and the objective of the scenario. During the scenario, the assessors are focused on observing the trainee. After the scenario, a debriefing takes place. Self-assessment questions can be asked during the debriefing:

- What went well?
- What went wrong?
- What would you do differently next time?

Based on the answers to these questions, the assessor gives the trainees indirect advice. Furthermore, during the debriefing of the scenario, some time should be spent on the critical criteria items to make the trainees aware of the importance of these items (why are these the critical items?) and to give them advice on how to deal or develop these soft skills. Instead of verbal, the trainee can also be asked to complete the assessment form him or herself as a form of self-assessment. In this case, it is important to discuss the differences with the observations of the assessor afterwards in a one-on-one meeting. During the scenario play, two assessors observe the trainee. They complete an observation form. At the end of the scenario they complete, independent, the rubric. After this, the two assessors discuss together their completed rubric and translate these scores into one final rubric. This will be done for at least two assessment scenarios.

Phase 4: Evaluation

After the last assessment scenario, the end result of the assessment should be discussed one on one with the trainee. The end result should consist of the final grade (1 – 5) and a brief explanation and argumentation. It is also important to repeat what will happen with the result, in line with what is explained during the introduction of the assessment. Furthermore, the trainee should get the opportunity to share his experience, ideas and improvements.

6 DISCUSSIONS

6.1 Limitations

Reliability and validity are important aspects of research. Reliability in qualitative research is challenging. Some researchers have critic on case study because of their limited ability to reliable generalization of the findings because of the small samples and using qualitative research. Since qualitative and mixed methods research is more valued, this criticism is losing favour and been countered by researchers (Flyvberg, 2011). According to Saunders et al. (2019) “purposive sampling can’t be considered to be statistically representative of the target population”. This case study is focused on a specific inland shipping company. Based on aspects as the maritime culture, the organizational culture, the soft skills needed for the officers and captain functions, the interviewees stated that this research can be used as a reference for other inland shipping companies. Nevertheless, when developing and using a plan of approach (for assessment and grading) it is always needed to specify the interests and needs involved in the organization.

Another limitation of this research is the use of online interviews and a survey for the trainees due to the pandemic. Most interview meetings are held online, via video tools and technology. During online meetings, non-verbal communication is partly missing and therefore indications of discomfort or enthusiasm can be missed. The interviews with the trainees are transformed into a survey, where interviews were preferred. But because of practical reasons and the pandemic, this was not possible. Therefore, the advantage of an interview to ask further on specific topics are missing.

6.2 Practical and ethical issues

During this research, the ethical aspect is always be taken into consideration. Informed consent and guarantee of confidentiality and anonymity must be ensured. The purpose of the research is explained to the interviewees and respondents, and they should be told that they have the right to choose whether to take part in the study or not (Cohen et al., 2001). All participants are provided with information concerning the research, including what it is about and any benefits it will provide. Anonymity is provided by informing the trainees that their identity will not be exposed in any way.

The current situation of the pandemic and the uncertainty because of this, have an impact on gathering data needed for this research. The economic impact of the pandemic on the maritime industry have influenced this research because there was no budget anymore for external training providers and simulator training. Another consequence of the pandemic is the change of the interviews with the trainees to a survey.

6.3 Suggestions for future research

During the study due to pandemic constraints, it was not possible to implement the plan of approach during an assessment and evaluate the plan of approach. There is no training budget at this moment because of the economic impact of the pandemic. Implementation and evaluation of this plan should be done in the future together with the three main stakeholders. Based on the evaluation missing items or additions can be added and/or further improvement of the plan of approach can be made.

7 CONCLUSION

This thesis studied answered the main question: *how to assess and grade soft skills in inland navigation that meets the transparency requests, interests and needs of the three main stakeholders?* To answer this main research question, a qualitative study was conducted. In order to give a framework for the main question, first, the conclusions on the sub-questions will be given.

A. What is the best environment for adult learners for learning and assessment?

The best environment for adult learners, the trainees, is a learner-centred approach that fits the best in combination with a safe and comfortable environment, with realistic and task-centred scenarios. It is also advised to use practical examples and to spend time on changing experience for interaction and two way learning. This is inspired by the andragogy theory. The experiential learning cycle of Kolb to create knowledge through the transformation of experience should also be included in learning and assessment.

B. What are soft skills?

Soft skills are seen as strategic elements in both private and professional life. Soft skills depend on the context and it is possible to improve them by training and education. They are not directly related to one position, but mainly about the relationship with other people. In the maritime sector, communication and cultural awareness are two examples of important soft skills.

C. What are methods to assess soft skills?

For the assessment of soft skills, the traditional way of assessment doesn’t fit. Several authors (Miller, 1990; Kobayashi, 2005) see simulators as a technical tool to show competencies and to perform a task. This is supported by the interviewees, they see simulators as the best tool to assess and grade

soft skills. As an important addition to the theory, all interviewees mentioned the use of at least two qualified assessors. Another addition is the importance and the time needed for an adequate briefing and debriefing when using a simulator. Without a briefing and debriefing, the use of a simulator is useless.

D. What are the criteria for the assessment and grading of soft skills?

Key criteria for assessment are:

- An assessment should be objective, integrity and transparent;
- The objectives of the assessment should be clear and be communicated with all stakeholders.
- The assessment should consist of realistic scenarios and at least two familiarization scenarios are needed before the start of the assessment. Before a scenario exercise, a briefing is needed and after the scenario, a debriefing should take place;
- Concerning the assessors, the results show that at least two assessors should observe the trainees.
- For the assessment and grading, a rubric should be used with a maximum of 10 criteria. Each criterion should be clearly stated, brief, observable, statement of behaviour and written in language the trainees understand.
- The rubric should include five levels per criterion, to use a grading scale from 1 to 5. For each criterion, the five levels should be described.
- The assessor has the final responsibility of the assessment, adding a self-assessment will support the commitment to the assessment and improve self-awareness.

The idea about peer assessment is not supported by all stakeholders. Expected is that this is because of the independent culture in inland navigation and no experiences with peer assessment. To support joint responsibilities, it is suggested to ask a representative group of trainees and

involve them occasionally during the process and not during the whole process as suggested in theory.

Grading of the assessments should be:

- Reliable
- Fairness
- Validity
- Clear boundaries and differences between the levels.

An overview of the conceptual framework, the authors and results based on the interviews and survey are represented in table 6.

TABLE 6. Conceptual framework, authors and results. (Bosker, 2021)

	Building stone framework	**Author**		**Results**
Assessment design	Authentic assessment toolbox: 4 steps for assessment development	Mueller (2020)	✓	The 4 steps are recognized by the interviewees.
Self-directed learning approach	Involvement of trainees in the design phase	Crowe (2000)		Too complex; a representative group is suggested.
Assessors	Facilitator, self and peer assessment	Crowe (2000); Hammond & Collins (1991)		Facilitator and self-assessment are supported.

Method for assessing soft skills	Use of simulator	Miller (1990); Kobayashi (2005)	✓	The best method to use for soft skills assessment.
Length of scenario	40 – 60 minutes	Kobayashi (2005)	✓	45-60 minutes for a scenario, including briefing and debriefing 1,5 – 2 hours.
Scoring scale of the performance	+1 (attained), 0 (lacked) and if applicable/possible -1 (missed).	Kobayashi (2005)		Most interviewees and trainees prefer to use a scoring scale of 1 (poor) up to 5 (excellent).
Grading criteria	Reliable and fairness	Felder & Brent (2010)	✓	Reliable and fair are seen as important criteria, validity and clear boundaries/differences are added.

E. How to implement the criteria to become an integral part of the training process?

The implementation of the assessment and grading system in the training process should go parallel to the design, implementation and evaluation process of a training course. Therefore it is required to know before the start of the design process what is expected by the customer, in this case the inland shipping company. What is the objective of the training process? Is an assessment desired, needed or required? And what is the objective of the assessment? These questions are needed to be answered during the preparation phase. In the preparation phase the criteria of the assessment are determined to be able to become an integral part during the design and execution phase.

F. What is the impact and value of this (soft skills) assessment and grading criteria to the maritime industry?

The impact of these soft skills and grading system approach for inland navigation is very valuable. For inland navigation, no study or research is available for specific training and assessment. This plan of approach for assessment and grading is the first research available for inland navigation. Although this study consists of a case study of one inland shipping company, the experts and external training provider have experience with assessments in maritime shipping and with training in inland navigation. Both training courses and assessments should be tailor-made and company-specific developed, however, the framework and criteria can be used for inland navigation assessment in general.

In conclusion, based on the results of:

- The literature;
- The interviews with the inland shipping company, the external training provider and experts;
- The survey from the trainees.

conducted in this study, an assessment and grading plan of approach is developed together will all three stakeholders and independent experts to answer the main question of this study. Conducting the interviews was very valuable since additions to the theory could be added and experiences could be related to the findings based on the literature. Also, the personal observations of the respondents gave insight into the expectations and threw some extra (practical) focus points on certain aspects within the training and assessment. Questioning the various stakeholders resulted in different insights, not opposite but always additional to each other since they have their own expertise. By questioning the various stakeholders, different complementary views and expertise came together, like the pieces of a puzzle that fall into place.

Based on the results, 4 phases are elaborated: preparation, design, execution and evaluation. The quality of the assessment depends on the combination and integration of realistic simulator scenarios, including briefing and debriefing, qualified assessors and a clear and understandable rubric with a grading system that distinguish 5 levels. In addition, the trainees should be instructed before the start of the assessment about the objective of the assessment, what will happen during the assessment and with the results of the assessment.

LIST OF REFERENCES

Albino, J., Young, S., Neumann, L., Kramer, G., Andrieu, S., Henson, L., Horn, B., Hendricson, W. (2009). Assessing Dental Students' Competence: Best Practice Recommendations in the Performance Assessment Literature and Investigation of Current Practices in Predoctoral Dental Education. Journal of dental education. 72. 1405-35. 10.1002/j.0022-0337.2008.72.12.tb04620.x.

Boashash, B. and Sucic, V. (2003), "Resolution measure criteria for the objective assessment of the performance of quadratic time-frequency distributions," in IEEE Transactions on Signal Processing, vol. 51, no. 5, pp. 1253-1263, doi: 10.1109/TSP.2003.810300

Bryman, A. (2016). Social research methods. UK: Oxford University Press.

Carson-Jackson, J. (2010). A simulation instructor's handbook, The Learning Game. London, The Nautical Institute.

Cimatti, B. (2016). Definition, Development, Assessment of soft skills and their role for the quality of organizations and enterprises. International Journal for Quality Research, 10(1), 97–130. https://doi.org/10.18421/IJQR10.01-05

Crowe, J. (2000). Evaluation of Adult Learners: Ethical Issues. New Horizons in Adult Education and Human Resource Development, 14(3), 4–10. https://doi.org/10.1002/nha3.10117

Dietel, R.J., Herman, J.L. & Knuth, R.A. (1991) What does research say about assessment? Oak Brook: NCREL.

Dubois, A. and Gadde, L E. (2002) 'Systematic combining: An abductive approach to case research', Journal of Business Research, Vol. 55, pp. 553– 60.

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), 19-26.

Erceg, B. Č. (2018). Legal Framework of European Inland Waterways and Croatian Legislation on Inland Waterways Navigation – Problems of Non-Harmonized Rules. European Scientific Journal, ESJ, 14(6). https://doi.org/10.19044/esj.2018.v14n6p%p

Felder, R. M., & Brent, R. (2010). Hard assessment of soft skills. Chemical Engineering Education, 44(1), 63-64.

Flyvberg, B. (2011) 'Case study', in N.K. Denzin and Y.S. Lincoln (eds) The Sage Handbook of Qualitative Research (4th edn). London: Sage, pp. 301– 16.

Hammond, M. & Collins, R. (1991). Self-directed learning: Critical practice. New York, NY: Nichols/GP Publishing.

Hanzu-Pazara, R., Barsan., E., Arsenie, P., Chiotoriou, L., Raicu, G. (2008). Reducing of maritime accidents caused by human factors using simulators in training process, Journal of Maritime Research, Vol. V, No 1, pp. 3-18

Houde, J. (2006). Andragogy and Motivation: An Examination of the Principles of Andragogy through Two Motivation Theories. Online Submission.

Illeris, Knud (2007): How We Learn: Learning and Non-learning in Schools and Beyond. London/ New York: Routledge.

Illeris, K. (Ed.). 2009. Contemporary Theories of Learning: Learning Theorists ... In Their Own Words (1 edition). London; New York: Routledge.

Isaac-Savage, P. (2009). Assessing adult learning and learning styles. Chapter 9: Assessing adult learning and learning styles. IGI Global. 10.4018/978-1-61520-745-9.

Knowles, M., Holton, E., & Swanson, R. (2005). The adult learner the definitive classic in adult education and human resource development (6th ed.). Elsevier.

Kolb, D. A. (1984). Experiential learning: Experience as the source of learning and development. Englewood Cliffs, NJ: Prentice-Hall.

Kolb, D. (2014). Experiential Learning: Experience as the Source of Learning and Development, 2nd edition. Pearson FT Press.

Kobayashi, H. (2005). Use of Simulators in Assessment, Learning and Teaching of Mariners. WMU Journal of Maritime Affairs, 4(1), 57–75. https://doi.org/10.1007/BF03195064

Maran, N. J. & Glavin, R., 2003, Low- to high-fidelity simulation: a continuum of medical education? Medical Education 37:22–28. https://doi.org/10.1046/j.1365-2923.37.s1.9.x

Maung, C. T. (2019), Simulation Training and Assessment System on Maritime Education and Training, World Maritime University.

Melnikovas, A. (2018), Towards an Explicit Research Methodology: Adapting Research Onion Model for Future Studies, Journal of Future Studies, 23 (2): 29-44.

Mueller, J. (2005). The Authentic Assessment Toolbox: Enhancing Student Learning through Online Faculty Development. Journal of Online Learning and Teaching. Volume 1, Number 1. http://jolt.merlot.org/documents/vol1_no1_mueller_001.pdf

Mueller, J. (2020), Authentic Assessment Toolbox website, accessed 23 August 2020. http://jfmueller.faculty.noctrl.edu/toolbox/index.htm

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. Academic medicine, 65(9), S63-7.

OECD (2002), Glossary of statistical terms: Inland waterways transport vessels, accessed 2 February 2021, https://stats.oecd.org/glossary/detail.asp?ID=4127

Pham, D. C. (2019). Improving simulation-based training to better serve the maritime community: a comparative research between the aviation and maritime domains.

Pinto, H., Cruz, A. R., & Combe, C. (2015). Cooperation and the emergence of maritime clusters in the Atlantic: Analysis and implications of innovation and human capital for blue growth. Marine Policy, 57, 167-177.

Policy Research Corporation (2008)The role of maritime clusters to enhance the strength and development of maritime sectors. European Commission.

Ramm, T., Thomson, A., Jackson, A. (2015). Learning clinical skills in the simulation suite: the lived experiences of student nurses involved in peer teaching and peer assessment. Nurse Education Today, 35(6), 823–827. https://doi.org/10.1016/j.nedt.2015.01.023

Reeves, T. C. (2000), Alternative assessment approaches for online learning environments in higher education, J. Educational Computing Research, Vol. 23(1) 101-111

Salman, A. K. M. D. W. (2013), The importance of using ship bridge simulator training to enhance the compentency of masters and watch-officers: a case study of the Iraqi dredging fleet, World Maritime University Dissertations.

Saunders, M. N. K., Thornhill, A., Lewis, P. (2019). Research Methods for Business Students, Pearson Education. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/tampere/detail.action?docID=5774742.

Schulz, B. (2008). The importance of soft skills: Education beyond academic knowledge. Journal of Lan-guage & Communication, 2(1), 146-154.

Sellberg, C. (2017). Simulators in bridge operations training and assessment: a systematic review and qualitative synthesis. WMU J Marit Affairs 16, 247–263. https://doi.org/10.1007/s13437-016-0114-8

Sellberg, C. (2018). From briefing, through scenario, to debriefing: the maritime instructor's work during simulator-based training. Cogn Tech Work (2018) 20:49–62. https://doi.org/10.1007/s10111-017-0446-y

Spalding, D. (2014). How to teach adults : plan your class, teach your students, change the world (Expanded edition.). Jossey-Bass.

Thomas, M. and Magilvy, J. K. (2011). Qualitative Rigor or Research Validity in Qualitative Research. Journal for Specialists in Pediatric Nursing, 16(2), 151–155. https://doi.org/10.1111/j.1744-6155.2011.00283.x

Whitehurst, G. J. R. (2016). Grading soft skills: The Brookings soft skills report card. Evidence Speaks Reports, 2(4), 1-7.

Yale University (2015), Fundamentals of Qualitative Research Methods: Interviews (module 3), Watched on 20.12.2020. https://www.youtube.com/watch?v=6PhcglOGFg8

Yin, R.K. (2018) Case Study Research and Applications: Design and Methods (6th edn). London: Sage.

APPENDICES

Appendix 1. Interview questions

Key principles:	
Experience	Can you tell me about your experience with assessment and grading of soft skills?
Main challenges	What are according to you the main challenges of soft skills assessment?
Requirements	Which requirements are according to you crucial for soft skills assessment and grading?
Breadth and balance of assessment design:	
Learner-centred learning approach	Some studies advice to involve the trainees into the whole process. Do you see opportunities to include trainees in the development phase? Can you provide an example?
Simulator	What methods and tools do you see as appropriate to assess and grade soft skills?
	Have you experience with simulator training and assessment? How did you experience this?

Duration	What is according to you the optimal length of the assessment scenario?
Execution of assessment:	
Joint responsibility / Self-directed learning approach	What would you say of including a combination of self, peer and facilitator assessment?
Rubric	The literature shows that rubrics can be used to assess performance along a task-specific set of criteria. What is your experience with using a rubric for assessment?
Evaluation of performance:	
Grading system	There are various grading systems possible (1-5; 0, 1 & 2; -1, 0, 1) Do you have experience with an adequate grading system? Which grading system do you think/experience as most supportive?
Criteria	The literature shows reliability and fair as crucial grading criteria. Do you agree with this? Do you miss an important criteria?

Appendix 2. Survey questions

1. What do you need to feel safe and comfortable?

During a training course or an assessment it is important to feel yourself safe and comfortable to be able to show your competencies. What do you need in this situation?

2. What information do you need in a briefing BEFORE the start of the assessment (exercise)?

Think about the (simulator) environment, documents, preparations. Information from the company and/or assessor.

3. What do you expect from the assessor?

4. What do you expect from a simulator exercise used for assessment?

5. What do you expect and need AFTER the assessment (exercise)?

6. To what extent, on a scale from 0-10, do you value the contribution of the assessment by an assessor, yourself and your peer(s) to your total assessment result?

A self-assessment is an individual review performed to identify elements that can be improved or exploited to achieve certain predefined assessment criteria. Peer assessment involves colleagues/candidates/students taking responsibility for assessing the work of their peers against set assessment criteria.

7. Which grading system do you prefer?

Select one answer

- Grading from 0 up to 3: where 0 is 'not observed', 1 is 'poor' and 3 is 'good'
- Grading from 0 up to 5: where 0 is 'not observed', 1 is 'poor', 3 is 'sufficient' and 5 is 'excellent'
- Grading by using - and/or + : -- / - / +/- / + / ++
- Grading by using -1 'poor', 0 'sufficient' and 1 'good'

- Other...

8. To what extent, on a scale from 0-10, do you experience the following elements as an crucial grading criteria?

- Reliability

- Fairness

- Validity

9. Anything you would like to add?